ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ANALYTIC THEORY OF SINGULAR ELLIPTIC PARTIAL DIFFERENTIAL EQUATIONS

BY W. J. TRJITZINSKY

(Received November 4, 1940)

1. Introduction

In this work we study the elliptic partial differential equation

$$F(u) \equiv \sum_{i,k=1}^{m} A_{i,k} \frac{\partial^2 u}{\partial x_i \partial x_k} + \sum_{i=1}^{m} B_i \frac{\partial u}{\partial x_i} + Cu = f \tag{1.1}$$

($A_{i,k} = A_{k,i}$; $m > 2$), where the $A_{i,k}$, B_i, C, f are functions of $(x) = (x_1, \cdots x_m)$, continuous for (x) in an open domain D in the Euclidean m-space of (x). Moreover,

$$D_1 A_{i,k}, \qquad D_2 A_{i,k}, \qquad D_1 B_i \tag{1.2}$$

are defined in D. In this connection, D_ν is a generic symbol for any of the partial derivatives of order ν.

It will be assumed that $f \subset L_2$ in D. Throughout this work any statement to the effect that a complex valued function $q \subset L_2$ in a domain R will be in the sense that q is measurable and that $|q|^2$ is integrable over R.

More precise hypotheses will be introduced in the sequel (see hypothesis (2.13)).

The type of problems, relating to (1.1), which we shall study are suggested, in part, by T. Carleman's[1] investigation of a Schrödinger wave equation, which in the sequel will be referred to as (C). Carleman utilizes for this purpose the methods of his theory of singular integral equations[2] in the sequel referred to as (C_1).

The case treated in (C) is

$$m = 3, \quad B_i = 0, \quad A_{i,k} = 0 \ (i \neq k), \quad A_{i,i} = 1 \qquad (i, k = 1, 2, 3). \tag{1.3}$$

We shall transform (1.1) into an integral equation. In view of our purposes our method, in this connection, will deviate from the customary ones. The transformation will be effected with the aid of certain developments which can be found in a book of J. Hadamard,[3] in the sequel referred to as (H).

---

[1] T. Carleman, *Sur la théorie mathématique de l'équation de Schrödinger* [Arkiv för Mat. Astron. och Fysik; 24B, Nr. 11 (1934), 1–7].

[2] T. Carleman, *Sur les équations intégrales singulières à noyau réel et symétrique* [Uppsala, 1923].

[3] J. Hadamard, *Le problème de Cauchy et les équations aux dérivees partielles linéaires hyperboliques* [Paris, 1932].

A considerable part of the developments in this work refers to equations containing a parameter λ, entering linearly in C. The results for λ non real and F self adjoint are partially analogous to those found in (C). The theory for λ real presents additional complications and has some analogy with a work of W. J. Trjitzinsky[4] in the field of singular integral equations.

In sections 2, 3 we carry out the transformation into an integral equation (Theorem 3.1). In section 4 the kernel of this equation is investigated in some detail (Theorem 4.1) and the equation is iterated a finite number of times, which yields an integral equation satisfying certain regularity properties. This equation plays an essential role in the subsequent developments. In section 5 we introduce characteristic functions and values (cf. (5.6), (5.9)) for certain approximating homogeneous boundary value problems ((5.7), (5.10)); we establish for these functions orthogonality properties (Theorem 5.1).

The developments in sections 6, 7, 8, 9, 10 *are for F self adjoint.*

In section 6 we first establish existence of solutions for the non homogeneous problem (5.1) when $\Im\lambda \neq 0$ (Theorem 6.1). In the complex λ-plane we define sets T (Definition 6.1) and particular sets T, which we term O (see text after (6.14)); sets T contain all the points off the axis of reals and may contain certain points on the axis of reals, depending on sufficient 'rarefication' of the set of points represented by the totality of characteristic values formed for an infinite sequence of approximating homogeneous boundary value problems. We establish existence of solutions, $\subset L_2$, of (5.1) *when* λ *is in* T (Theorem 6.2). In section 7 there is developed a corresponding spectral theory (Theorem 7.1). In sections 8 and 9 this theory is applied; we introduce the notion of 'closure' of F with respect to a spectrum θ (Definition 8.1) and give a generalized Fourier expansion of functions, $\subset L_2$ (in D), in terms of θ (Theorem 8.1). Further, there is given a representation for solutions, referred to in Theorem 6.2, with the aid of θ (Theorem 8.2). In section 9 various representations, convergent in the mean square, are given. In section 10 we introduce certain operators $A_x(\lambda \mid \cdots)$, with the aid of which certain particular solutions of (5.1) may be expressed for λ in O (Theorem 10.1). With the aid of these operators we investigate conditions under which F is of class I (Theorem 10.2, Definition 6.2) and, finally, questions of uniqueness are studied for λ non real (Theorem 10.3) and for λ in O (Theorem 10.4).

In section 5 a certain "regularity" hypothesis is introduced for the domain D; this is needed only in the developments subsequent to (5.7). Some lightening of the conditions imposed on the coefficients in (1.1) is possible without any essential change in the conclusions; for simplicity this lightening of the hypothesis will be avoided.

The case when F is of class I corresponds (in the field of partial differential equations), as pointed out by Carleman, to the case of "Grenzpunkt" in the well known theory of H. Weyl in the field of ordinary differential equations.

---

[4] W. J. Trjitzinsky, *Some problems in the theory of singular integral equations* [Annals of Math. (1940), 584–619].

Every solution u of (3.37), such that the D_2u are continuous in D, is solution of (1.1) in the ordinary sense; (3.37) thus generalizes the operator F; *with F assigned this generalized meaning, the partial differential equation and the integral equation are equivalent.*

2. Preliminaries

The adjoint of $F(u)$ is

$$G(v) = \sum_{i,k} \frac{\partial^2}{\partial x_i \partial x_k}(A_{ik}v) - \sum_i \frac{\partial}{\partial x_i}(B_i v) + Cv. \tag{2.1}$$

The characteristic form for (1.1) is

$$A(\gamma_1, \cdots \gamma_m) = \sum_{i,k} A_{i,k}\gamma_i\gamma_k = A(\gamma_1, \cdots \gamma_m; x_1, \cdots x_m). \tag{2.2}$$

We shall make use of the symbol

$$\frac{\partial}{\partial \nu}$$

denoting operation of derivation along the *transversal direction* (cf. (H; 87)). This operation is to be understood as follows. We form

$$A = A(\pi_1, \cdots \pi_m), \tag{2.3}$$

where the π_i are direction cosines of normal to a surface, say

$$H \equiv H(x_1, \cdots x_m) = 0;$$

that is

$$\pi_i = \pm\left[\left(\frac{\partial H}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial H}{\partial x_m}\right)^2\right]^{-\frac{1}{2}} \frac{\partial H}{\partial x_i}, \tag{2.4}$$

the supposition being, of course, that the D_1H exist in the domain under consideration. The transversal direction is defined by the ordinary differential system (generally non linear):

$$\frac{dx_1}{w_1} = \frac{dx_2}{w_2} = \cdots = \frac{dx_m}{w_m} = d\nu, \tag{2.5}$$

where

$$w_k = w_k(x_1, \cdots x_m) = \sum_{i=1}^{m} A_{i,k}\pi_i \qquad (k = 1, \cdots m).^5 \tag{2.5a}$$

In the case when

$$\sum_{i,k} A_{i,k} \frac{\partial^2}{\partial x_i \partial x_k} = \Delta \qquad \text{(Laplacian)}$$

[5] See (H; 88) for a detailed discussion of the transversal direction.

the "transversal" direction will signify "normal" direction. In view of (2.5), (2.5a) we are brought to the definition

$$\frac{\partial u}{\partial \nu} = \sum_{k=1}^{m} \frac{\partial u}{\partial x_k}\frac{dx_k}{d\nu} = \sum_{k=1}^{m} \frac{1}{2}\frac{\partial A}{\partial \pi_k}\frac{\partial u}{\partial x_k} \qquad \text{(cf. (2.3))}. \tag{2.6}$$

There is on hand a following fundamental formula (see (H))

$$\int^{(m)} [vF(u) - uG(v)]\,dt = -\int^{(m-1)} \left[v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu} + Luv\right] dS. \tag{2.7}$$

[F from (1, 1); G from (2.1)]. Here the first integral displayed is over the space T bounded by the surface $H = 0$; the integral of the second member is over the surface S constituting the frontier of T;[6] moreover,

$$L = \sum_i \pi_i B_i - \sum_{i,k} \pi_i \frac{\partial A_{i,k}}{\partial x_k} \qquad \text{(cf. (2.4))} \tag{2.7a}$$

and

$$dT = dx_1 \cdots dx_m, \qquad dS = \frac{\partial H}{\partial n}\frac{dT}{dH} = \left[\left(\frac{\partial H}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial H}{\partial x_m}\right)^2\right]^{\frac{1}{2}} \frac{dT}{dH}.$$

Of importance for us will be a function of (x) and (z),

$$\Gamma(x, z) = \Gamma(x_1, \cdots x_m\,; z_1, \cdots z_m),$$

satisfying the partial differential equation

$$A\left(\frac{\partial \Gamma}{\partial x_1}, \cdots \frac{\partial \Gamma}{\partial x_m}; x_1, \cdots x_m\right) = 4\Gamma; \tag{2.8}$$

namely, the function for which

$$\Gamma^{\frac{1}{2}}(x; z) = r(x; z)$$

denotes the "*geodesic distance*"[3] between the points (x) and (z). This distance is formed with respect to the quadratic form

$$H = H(\gamma_1, \cdots \gamma_m) = \sum_{i,k} H_{i,k}(x)\gamma_i\gamma_k, \tag{2.9}$$

"*associated*" with the quadratic form $A(\gamma_1, \cdots \gamma_m)$ (2.2). In this connection it is to be noted that, if

$$(A_{i,j}) \qquad (i, j = 1, \cdots m)$$

denoted a matrix with $A_{i,j}$ in the i-th row and j-th column, then

$$(H_{i,j}) = (A_{i,j})^{-1};$$

that is, the matrix $(H_{i,j})$ is the inverse of the matrix $(A_{i,j})$.

With the aid of the developments given in (H; pp. 120–124) we shall study $\Gamma(x; z)$ in greater detail.

---

[6] This is asserted under suitable regularity conditions, in the sequel satisfied.

Consider the differential problem

$$\frac{dx_i}{ds} = \frac{1}{2}\frac{\partial A}{\partial p_i}, \qquad \frac{dp_i}{ds} = -\frac{1}{2}\frac{\partial A}{\partial x_i} \qquad (i = 1, \cdots m), \tag{2.10}$$

where

$$A = \sum_{i,k} A_{i,k}(x) p_i p_k, \tag{2.11}$$

while the initial conditions are

$$x_i = z_i \quad (\text{for } s = 0), \qquad p_i = p_{0,i} \quad (\text{for } s = 0); \tag{2.11a}$$

we write

$$A_0(p_0) \equiv \sum_{i,k} A_{i,k}(z) p_{0,i} p_{0,k} = c.$$

As indicated in (H; pp. 120, 121) the solution of this problem may be written in the form

$$x_i = \varphi_i(q_1, \cdots q_m ; z_1, \cdots z_m) \qquad (q_i = s p_{0,i}), \tag{2.12}$$

$$P_i = \psi_i(q_1, \cdots q_m ; z_1, \cdots z_m) \qquad (P_i = s p_i);$$

from these relations one obtains

$$z_i = \varphi_i(-P_1, \cdots, -P_m ; x_1, \cdots x_m), \quad q_i = \psi_i(-P_1, \cdots, -P_m ; x_1, \cdots x_m).$$

It is observed that (2.10) determines geodesics, with respect to the form H, through the point (z). Moreover, for solutions of (2.10) one has $A(p, x) =$ constant. On substituting $s = 0$ and using the initial conditions this constant is found to be c.

In consequence of known theorems for ordinary differential equations[7] we infer the following. If the

$$A_{i,k}(x), \qquad D_\nu A_{i,k}(x) \qquad (\nu = 1, \cdots 4) \tag{2.13}$$

are continuous for (x) in D, *which we henceforth assume*, then the elements x_i, P_i of the solution (2.12) are continuous and have continuous partial derivatives of the first and second orders with respect to the q_j and the z_j; furthermore, the third order derivatives will exist. More generally, if the $D_{N+1}A_{i,k}(x)$ are continuous, the partial derivatives with respect to the q_j and z_j, of order $N - 1$, of the x_i and P_i will be continuous and the derivatives of order N will exist. The above statements refer to local properties—for (z) representing a point in D.

According to (H)

$$\Gamma(x, z) \equiv A(q_1, \cdots q_m ; z_1, \cdots z_m) = \sum_{i,k} A_{i,k}(z) q_i q_k. \tag{2.14}$$

---

[7] Cotton, *Sur les équations différentielles dépendant de paramètres arbitraires* [Bull. Soc. mat., t. 37 (1909), 204–214]. Also see t. 38 (1918), p. 4.

Here the q_i are to be replaced by the expressions, in terms of the z_j and x_j, obtainable from the first m relations (2.12).

We have

$$x_i = x_i(s) = x_i(0) + x_i^{(1)}(0)\frac{s}{1!} + (x_i^{(2)}(0) + \xi_i(s))\frac{s^2}{2!}, \tag{2.15}$$

where

$$\lim_{s\to 0} \xi_i(s) = 0 \qquad (i = 1, \cdots m).$$

By (2.11a) $x_i(0) = z_i$; on the other hand, in view of (2.10) and (2.11)

$$x_i^{(1)}(0) = \frac{1}{2}\frac{\partial A}{\partial p_i}(s = 0) = \sum_k A_{i,k}(x)p_k(s = 0) = \sum_k A_{i,k}(z)p_{0,k}$$

so that, by definition of q_i,

$$sx_i^{(1)}(0) = \sum_k A_{i,k}(z)q_k . \tag{2.16}$$

By (2.10)

$$x_i^{(2)}(s) = \frac{d}{ds}\sum_\nu A_{i,\nu}(x)p_\nu = \sum_\nu A_{i,\nu}(x)\frac{dp_\nu}{ds} + \sum_\nu p_\nu \frac{dA_{i,\nu}(x)}{ds}. \tag{2.17}$$

Now the first sum in the last member of (2.17) is

$$g(s) = -\frac{1}{2}\sum_\nu A_{i,\nu}(x)\frac{\partial A}{\partial x_\nu} = -\frac{1}{2}\sum_{\nu,\alpha,\beta} A_{i,\nu}(x)\frac{\partial A_{\alpha,\beta}(x)}{\partial x_\nu}p_\alpha p_\beta$$

moreover, since $q_j = sp_{0,j}$,

$$s^2 g(0) = -\frac{1}{2}\sum_{\nu,\alpha,\beta} A_{i,\nu}(z)\frac{\partial A_{\alpha,\beta}(z)}{\partial z_\nu}q_\alpha q_\beta . \tag{2.17a}$$

We have

$$\frac{dA_{i,\nu}(x)}{ds} = \sum_{\gamma=1}^{m}\frac{\partial A_{i,\nu}(x)}{\partial x_\gamma}\frac{dx_\gamma}{ds};$$

whence by (2.16)

$$\frac{dA_{i,\nu}(x)}{ds}\Big]_{s=0} s = \sum_{\gamma=1}^{m}\frac{\partial A_{i,\nu}(z)}{\partial z_\gamma}sx_\gamma^{(1)}(0) = \sum_{\gamma,k} A_{\gamma,k}(z)\frac{\partial A_{i,\nu}(z)}{\partial z_\gamma}q_k$$

and

$$\frac{1}{2}\sum_\nu p_\nu \frac{dA_{i,\nu}(x)}{ds}\Big]_{s=0} s^2 = \sum_{\nu,\gamma,k}\frac{1}{2}A_{\gamma,k}(z)\frac{\partial A_{i,\nu}(z)}{\partial z_\gamma}q_k q_\nu . \tag{2.17b}$$

In consequence of (2.17a), (2.17b) from (2.17) one infers that

$$\frac{s^2}{2!}x_i^{(2)}(0) = \frac{s^2}{2}g(0) + \text{last member of (2.17b)} = \sum_{\alpha,\beta} A_{i:\alpha,\beta}(z)q_\alpha q_\beta,$$

$$A_{i:\alpha,\beta}(z) = -\frac{1}{4}\sum_\nu A_{i,\nu}(z)\frac{\partial A_{\alpha,\beta}(z)}{\partial z_\nu} + \frac{1}{2}\sum_\nu A_{\nu,\beta}(z)\frac{\partial A_{i,\alpha}(z)}{\partial z_\nu}. \tag{2.18}$$

Finally, *one may write* (2.15) *in the form*

$$x_i - z_i = \sum_k A_{i,k}(z)q_k + \sum_{\alpha,\beta} (A_{i:\alpha,\beta}(z) + \lambda_{i:\alpha,\beta}(z;\, q))q_\alpha q_\beta\,, \tag{2.19}$$

where the $A_{i:\alpha,\beta}(z)$ *are from* (2.18) *and*

$$\lim \lambda_{i:\alpha,\beta}(z;\, q) = 0 \qquad (\text{as } (q) \to (0);\ (z) \text{ in } D);$$

furthermore, the $\lambda_{i:\alpha,\beta}$ are continuous in

$$z_1\,, \cdots z_m\,, q_1\,, \cdots q_m \qquad (|\, q_j\,| \leqq q_0\,;\ q_0 > 0;\ (z) \text{ in } D).$$

Inversion of (2.19) will give

$$q_i = \sum_k H_{i,k}(z)(x_k - z_k) + \sum_{\alpha,\beta} (H_{i:\alpha,\beta}(z) + \bar{\lambda}_{i:\alpha,\beta})(x_\alpha - z_\alpha)(x_\beta - z_\beta), \tag{2.20}$$

where

$$H_{i:\sigma,\tau}(z) = -\sum_{\alpha,\beta,k} A_{k:\alpha,\beta}(z)H_{i,k}(z)H_{\alpha,\sigma}(z)H_{\beta,\tau}(z) \qquad (\text{cf. (2.18)}) \tag{2.20a}$$

and the $\bar{\lambda}$ are continuous in the x_j , z_j , while

$$\lim \bar{\lambda}_{i:\alpha,\beta} = \lim \bar{\lambda}_{i:\alpha,\beta}(z;\, x) = 0 \qquad (\text{as } (x) \to (z);\ (z) \text{ in } D). \tag{2.20b}$$

We note that the $H_{i:\sigma,\tau}$ are polynomials in the

$$A_{i,j}(z), \qquad D_1 A_{i,j}(z), \qquad H_{i,j}(z).$$

Accordingly, the functions of (2.20a) are continuous in D.

Substitution of (2.20) into (2.14) will yield

$$\Gamma(x,\, z) = \sum_{k,k_1} H_{k,k_1}(z)(x_k - z_k)(x_{k_1} - z_{k_1}) + 2 \sum_{\alpha,\beta,k_1} \bar{H}_{k_1:\alpha,\beta}(x,\, z)(x_\alpha - z_\alpha)(x_\beta - z_\beta)(x_{k_1} - z_{k_1}) + \Gamma_4\,,$$

where

$$\bar{H}_{k_1:\alpha,\beta}(x,\, z) = H_{k_1:\alpha,\beta}(z) + \bar{\lambda}_{k_1:\alpha,\beta} \qquad (\text{cf. (2.20a), (2.20b)})$$

and

$$\Gamma_4 = \sum_{\alpha,\beta,\alpha_1,\beta_1} (x_\alpha - z_\alpha)(x_\beta - z_\beta)(x_{\alpha_1} - z_{\alpha_1})(x_{\beta_1} - z_{\beta_1}) \sum_{r,s} A_{r,s}(z)\bar{H}_{r:\alpha,\beta}\bar{H}_{s:\alpha_1,\beta_1}\,.$$

Finally, *we obtain*

$$\Gamma(x,\, z) = \sum_{k,k_1} H_{k,k_1}(z)(x_k - z_k)(x_{k_1} - z_{k_1}) + \Gamma_3(x,\, z), \tag{2.21}$$

where

$$\Gamma_3(x,\, z) \equiv \sum_{\alpha,\beta,k_1} (2H_{k_1:\alpha,\beta}(z) + \lambda^*_{k_1:\alpha,\beta}(x,\, z))(x_\alpha - z_\alpha)(x_\beta - z_\beta)(x_{k_1} - z_{k_1}) \tag{2.21a}$$

(cf. (2.20a)); *here*

$$\lambda^*_{k_1:\alpha,\beta}(x,\, z) \to 0 \qquad (\text{as } (x) \to (z);\ (z) \text{ in } D); \tag{2.21b}$$

the functions, involved in (2.21b), *are continuous in* (x) *for* (x) *in a neighborhood of* (z).

When

$$\sum_{i,k} A_{i,k} \frac{\partial^2}{\partial x_i \partial x_k} = \Delta \qquad \text{(Laplacian)},$$

we have

$$\Gamma(x, z) = (x_1 - z_1)^2 + \cdots + (x_m - z_m)^2.$$

3. The transformation

In view of the definition (2.6) one has

$$\frac{\partial}{\partial \nu} \Gamma(x, z) = \sum_{k=1}^{m} \frac{1}{2} \frac{\partial A}{\partial \pi_k} \frac{\partial \Gamma}{\partial x_k} \qquad \text{(cf. (2.4), (2.3))}, \tag{3.1}$$

the derivative in the transversal direction being with respect to a surface $H = 0$. With $\rho(> 0)$ suitably small we consider the surface

$$H_\rho(x) \equiv \Gamma(x, z) - \rho^2 = 0 \qquad ((z) \text{ fixed in } D). \tag{3.2}$$

Correspondingly

$$\pi_i = \frac{\partial \Gamma}{\partial x_i} \frac{1}{\sigma(x, z)}, \qquad \sigma^2(x, z) = \left(\frac{\partial \Gamma}{\partial x_1}\right)^2 + \cdots + \left(\frac{\partial \Gamma}{\partial x_m}\right)^2. \tag{3.3}$$

By (2.3) and (3.1)

$$\frac{\partial}{\partial \nu} \Gamma(x, z) = \sum_{k=1}^{m} \sum_{\alpha} A_{\alpha,k}(x) \pi_\alpha \frac{\partial \Gamma}{\partial x_k}.$$

Thus, in view of (3.3),

$$\frac{\partial}{\partial \nu} \Gamma(x, z) = \sum_{\alpha,k} A_{\alpha,k}(x) \frac{\partial \Gamma}{\partial x_\alpha} \frac{\partial \Gamma}{\partial x_k} \frac{1}{\sigma(x, z)}$$

and, by virtue of (2.8),

$$\frac{\partial}{\partial \nu} \Gamma(x, z) = \frac{1}{\sigma(x, z)} 4\Gamma(x, z).$$

Whence, when α is a constant,

$$\frac{\partial}{\partial \nu} \Gamma^\alpha(x, z) = \alpha \Gamma^{\alpha-1}(x, z) \frac{\partial}{\partial \nu} \Gamma(x, z) = \frac{4\alpha}{\sigma(x, z)} \Gamma^\alpha(x, z) \qquad \text{(cf. (3.3))}. \tag{3.4}$$

With a view to applying the fundamental formula (2.7) we wish to find a function

$$v(x, z) = I(\Gamma(x, z)) \tag{3.5}$$

such that

$$v = \frac{\partial v}{\partial \nu} = 0 \qquad ((z) \text{ fixed in } D) \tag{3.5a}$$

for (x) on the surface $H_\rho(x) = 0$ (3.2).

On taking account of (3.4) it is observed that a function (3.5) satisfying (3.5a) may be taken of the form

$$v(x, z) = \Gamma^{\frac{1}{2}(2-m)}(x, z) + \frac{1}{\rho^{2m-4}}\,\Gamma^{\frac{1}{2}(m-2)}(x, z) - \frac{2}{\rho^{m-2}}. \tag{3.6}$$

Inasmuch as $m > 2$, it is noted that v becomes infinite essentially as $\Gamma^{1-\frac{1}{2}m}$, when $(x) \to (z)$.

With

$$r^2 = (x_1 - z_1)^2 + \cdots + (x_m - z_m)^2 \qquad ((z) \text{ fixed in } D) \tag{3.7}$$

the differential element of volume, dT, may be given in the form

$$dT = r^{m-1}\,dr\,d\Omega, \tag{3.7a}$$

where $d\Omega$ *is element of surface of a hypersphere of radius unity* in the m-space of (x). We recall that the total surface of this hypersphere is $S_m = 2\pi^{\frac{1}{2}m}/g\left(\frac{m}{2}\right)$, where $g(u)$ is the Gamma-function. The volume is $\frac{1}{m}\,S_m$. From (3.7a) it is inferred that the element of surface of a hypersphere of radius r is

$$dw = \frac{dT}{dr} = r^{m-1}\,d\Omega. \tag{3.7b}$$

The element dS of the surface (3.2) satisfies the relation

$$dS \cos \tau = dw, \tag{3.8}$$

where τ is the angle formed by the line extending from (z) and having direction cosines $(x_i - z_i)/r$ and by a suitable direction of the normal to the surface (3.2) at (x); the direction cosines of this normal are the π_i of (3.3) (with a suitable sign for $\sigma(x, z)$); thus,

$$\cos \tau = \frac{1}{r\sigma(x, z)} \sum_i \frac{\partial \Gamma(x, z)}{\partial x_i}\,(x_i - z_i)$$

and, by (3.8) and (3.7b),

$$dS = \frac{\sigma(x, z) r^m\,d\Omega}{\sum_i (x_i - z_i)\,\dfrac{\partial \Gamma(x, z)}{\partial x_i}} \qquad \text{(cf. (3.3))}. \tag{3.9}$$

Let K_δ denote a sphere with center at (z) and radius δ $(0 < \delta \leqq \delta_0)$, where δ_0 is taken sufficiently small so that the surface S_δ, of K_δ, lies interior the domain Γ_ρ bounded by the surface (3.2).[8] We have a domain

$$D(\rho, \delta), \tag{3.10}$$

whose frontier consists of the spherical surface S_δ and of the surface $H_\rho(x) = 0$.

[8] Γ_ρ contains (z) in the interior and is topologically a sphere.

Suppose $u = u(x)$ is a function, continuous with the D_1u, satisfying in D the differential equation (1.1). We substitute in the fundamental formula (2.7) u equal to the aforesaid function and $v = v(x, z)$, from (3.6), and we extend the integrations over the domain $D(\rho, \delta)$ (cf. (3.10)) and over the frontier of $D(\rho, \delta)$; the variable of integration is to be (x). Thus,

$$\int_{D(\rho,\delta)}^{(m)} [vf - uG(v)]\, dT = Q(\rho, \delta) \tag{3.10a}$$

where

$$Q(\rho, \delta) = -\int_{H_\rho + S_\delta}^{(m-1)} \left[v \frac{\partial u}{\partial \nu} - u \frac{\partial v}{\partial \nu} + Luv \right] dS.$$

On taking account of (3.5a) one obtains

$$Q(\rho, \delta) = -\int_{S_\delta}^{(m-1)} \left[v \frac{\partial u}{\partial \nu} - u \frac{\partial v}{\partial \nu} + Luv \right] dS. \tag{3.11}$$

With the notation (3.7) we have

$$r = \delta \qquad \text{(for } (x) \text{ on } S_\delta);$$

moreover, by virtue of the statement referring to (3.7b), dS in (3.11) is of the form

$$dS = dw(r = \delta) = \delta^{m-1} d\Omega. \tag{3.11a}$$

In (3.11) L is defined by (2.7a), where the π_i are the direction cosines of the normal to the surface S_δ, extended in the direction interior $D(\rho, \delta)$; that is, the normal in question is directed outwards from the sphere K_δ. We have

$$\pi_i = \frac{x_i - z_i}{\delta} \qquad ((x) \text{ in } S_\delta)$$

and

$$L = \sum_i B_i(x)\pi_i - \sum_{i,k} \frac{\partial A_{i,k}(x)}{\partial x_k} \pi_i. \tag{3.11b}$$

By definition (2.6)

$$\frac{\partial u}{\partial \nu} = \sum_{k=1}^{m} \frac{1}{2} \frac{\partial A}{\partial \pi_k} \frac{\partial u}{\partial x_k} = \sum_{k,i} A_{i,k}(x)\pi_i \frac{\partial u}{\partial x_k} \qquad \text{(on } S_\delta). \tag{3.21}$$

Now

$$\frac{\partial \Gamma}{\partial \nu} = \sum_{k,i} A_{i,k}(x)\pi_i \frac{\partial \Gamma}{\partial x_k} \tag{3.13}$$

and, in consequence of (3.6),

$$\frac{\partial v}{\partial \nu} = \left[\frac{2 - m}{2} \Gamma^{-\frac{1}{2}m}(x, z) + \frac{1}{\rho^{2m-4}} \left(\frac{m - 2}{2} \right) \Gamma^{\frac{1}{2}m-2}(x, z) \right] \frac{\partial \Gamma}{\partial \nu}. \tag{3.14}$$

Before proceeding further we shall study the function

$$w(x, z) = \frac{r}{\sqrt{\Gamma(x, z)}} \qquad (r^2 = (x_1 - z_1)^2 + \cdots + (x_m - z_m)^2). \tag{3.15}$$

By (2.21)

$$w^2(x, z) = [H(z, \pi) + g_3(x, z)]^{-1} \tag{3.16}$$

where

$$H(z, \pi) = \sum_{k,k_1} H_{k,k_1}(z)\pi_k \pi_{k_1}, \qquad \pi_k = \frac{x_k - z_k}{r} \tag{3.16a}$$

and

$$g_3(x, z) = \frac{1}{r^2}\Gamma_3(x, z) \qquad \text{(cf. (2.21a))}. \tag{3.16b}$$

For $r \leqq \delta_0$

$$| 2H_{k_1;\alpha,\beta}(z) + \lambda^*_{k_1;\alpha,\beta}(x, z) | \leqq \lambda_0(z), \tag{3.16c}$$

where $\lambda_0(z)$ is independent of (x). Accordingly, by (3.16b)

$$| g_3(x, z) | \leqq r \sum_{\alpha,\beta,k_1} \lambda_0(z)\pi_\alpha \pi_\beta \pi_{k_1} \leqq r\lambda(z) \qquad (0 \leqq r \leqq \delta_0) \tag{3.16d}$$

($\lambda(z)$ independent of (x)). Now $H(z, \pi)$ is a positive definite quadratic form; thus, inasmuch as $\pi_1^2 + \cdots + \pi_m^2 = 1$, one has

$$H(z, \pi) \geqq h(z) = \sum_{k,k_1} H_{k,k_1}(z)\pi'_k \pi'_{k'} > 0 \tag{3.17}$$

(for some values π'_k, with $\pi_1'^2 + \cdots + \pi_m'^2 = 1$), where $h(z)$ is independent of (π). Hence, in view of (3.17) and (3.16d)

$$H(z, \pi) + g_3(x, z) \geqq h(z) - r\lambda(z) \geqq h(z) - \delta_0\lambda(z) = h_0^2(z) > 0 \tag{3.17a}$$

for $r \leqq \delta_0$, provided one takes

$$0 < \delta_0 < \frac{h(z)}{\lambda(z)}. \tag{3.17b}$$

Whence, by (3.16)

$$| w(x, z) | \leqq \frac{1}{h_0(z)} < \infty \qquad (0 \leqq r \leqq \delta_0; \text{ cf. (3.15), (3.17a), (3.17b)}). \tag{3.18}$$

It is essential to note that *in* (3.18) $h_0(z)$ *is independent of* (x).

On writing

$$w^2(x, z) = \frac{1}{H(z, \pi)} + \rho_0(x, z),$$

by (3.16), (3.16d), (3.17) and (3.17a) we have

$$|\rho_0(x,z)| = \frac{|g_3(x,z)|}{|H(z,\pi)|\,|H(z,\pi)+g_3(x,z)|} \leqq \frac{r\lambda(z)}{h(z)h_0^2(z)}$$

for $0 \leqq r \leqq \delta_0$. Accordingly, *the function of* (3.15) *satisfies the relation*

$$w^2(x,z) = \frac{1}{H(z,\pi)} + r\rho(x,z) \qquad \text{(cf. (3.16a))} \tag{3.19}$$

where

$$|\rho(x,z)| \leqq \rho(z) = \frac{\lambda(z)}{h(z)h_0^2(z)} \qquad (r \leqq \delta_0); \tag{3.19a}$$

here $\rho(z)$ *is independent of* (x). In particular

$$\lim_{r\to 0} w(x,z) = H^{-\frac{1}{2}}(z,\pi), \tag{3.19b}$$

provided $(x) \to (z)$ along a fixed direction whose direction cosines are $\pi_1, \cdots \pi_m$. *In* (3.19b) *the limit is attained uniformly with respect to the* π_j; that is, with respect to direction of approach. Of course, the limit depends on the direction of approach, but there is uniformity in the sense implied by (3.19), (3.19a).

The function of (3.11) will be expressed in the form

$$Q(\rho,\delta) = Q_1(\rho,\delta) + Q_2(\rho,\delta), \tag{3.20}$$

where, on taking account of (3.11a),

$$Q_1(\rho,\delta) = -\int^{(m-1)} \left[\frac{\partial u}{\partial \nu} + Lu\right] v\delta^{m-1}\,d\Omega \qquad \text{(cf. (3.11b), (3.12))}, \tag{3.20a}$$

and

$$Q_2(\rho,\delta) = \int^{(m-1)} u\,\frac{\partial v}{\partial \nu}\,\delta^{m-1}\,d\Omega \qquad \text{(cf. (3.14), (3.13))}. \tag{3.20b}$$

If we denote by l the upper bound in the spherical domain K_δ of

$$\sum_i |B_i(x)| + \sum_{i,k}\left|\frac{\partial A_{i,k}(x)}{\partial x_k}\right|,$$

in consequence of (3.11b) it is deduced that

$$|L| \leqq l \qquad \text{(on } S_\delta\text{; for } 0 < \delta \leqq \delta_0\text{)}.$$

On the other hand, by (3.12), for (x) on S_δ one has

$$\left|\frac{\partial u}{\partial \nu}\right| \leqq \sum_{k,i} |A_{i,k}(x)|\left|\frac{\partial u}{\partial x_k}\right| \leqq l_1 \qquad (0 < \delta \leqq \delta_0), \tag{3.21}$$

where l_1 is independent of δ. In this connection one may define l_1 as the upper-bound of the second member in (3.21) for (x) in the spherical domain K_δ.[9] Similarly,

$$|u| \leqq l_2 \qquad \text{(on } S_\delta\text{; } l_2 \text{ independent of } \delta(\leqq \delta_0)).$$

In consequence of (3.6)

$$v\delta^{m-1} = \delta\left\{\left(\frac{\delta}{\sqrt{\Gamma}}\right)^{m-2} + f(x, z)\delta^{m-2}\right\} \qquad ((x) \text{ on } S_\delta), \tag{3.22}$$

where

$$f(x, z) = \frac{1}{\rho^{2m-4}}\Gamma^{\frac{1}{2}(m-2)}(x, z) - \frac{2}{\rho^{m-2}}.$$

Clearly

$$f(x, z) \leqq f_0(z) < \infty \qquad (0 \leqq r \leqq \delta_0),$$

$f_0(z)$ being independent of (x). Thus, by (3.22) and (3.18)

$$|v\delta^{m-1}| \leqq \delta\{h_0^{-m+2}(z) + f_0(z)\delta^{m-2}\} \leqq \delta f_1(z) \qquad ((x) \text{ on } S_\delta; 0 < \delta \leqq \delta_0) \tag{3.23}$$

with $f_1(z)$ $(< \infty)$ independent of (x) and δ.

On taking account of (3.21) and (3.23), from (3.20a) it is inferred that

$$|Q_1(\rho, \delta)| \leqq \delta \int^{(m-1)} (l_1 + ll_2)f_1(z)\, d\Omega = \delta(l_1 + ll_2)f_1(z)S_m;$$

here S_m is the area of the surface of a hypersphere (in the space of (x)) of radius unity. *Consequently*

$$\lim_{\delta \to 0} Q_1(\rho, \delta) = 0. \tag{3.24}$$

Turning our attention to the integral (3.20b) we investigate the integrand involved. In view of (2.21) and of certain other considerations of section 2 one has

$$\frac{\partial \Gamma}{\partial x_k} = \sum_{k_1} (2H_{k,k_1}(x) + \omega_{k,k_1}(z, x))(x_{k_1} - z_{k_1}),$$

where the ω are continuous in (z) and (x), while

$$\omega_{k,k_1}(z, x) \to 0 \qquad (\text{as } (x) \to (z)) \tag{3.25}$$

uniformly with respect to direction of approach. Whence by virtue of (3.13)

$$\begin{aligned}\frac{\partial \Gamma}{\partial \nu} &= \sum_{k,i} A_{i,k}(x)\pi_i \sum_{k_1} (2H_{k,k_1}(x) + \omega_{k,k_1}(z, x))(x_{k_1} - z_{k_1}) \\ &= \sum_{k,i,k_1} A_{i,k}(x)\omega_{k,k_1}(z, x)\pi_i(x_{k_1} - z_{k_1}) + \gamma,\end{aligned}$$

---

[9] It is implied that the $\dfrac{\partial u}{\partial x_k}$ are continuous in D.

where

$$\gamma = 2\sum_{i,k_1}\pi_i(x_{k_1}-z_{k_1})\sum_k A_{i,k}(x)H_{k,k_1}(x) = 2\sum_{i=1}^{m}\pi_i(x_i-z_i) = 2\delta.$$

Hence, the transversal derivative being with respect to the spherical surface S_δ,

$$\frac{1}{\delta}\frac{\partial\Gamma}{\partial\nu} = 2+\omega'(z,x) \qquad ((x) \text{ on } S_\delta);$$

here $\omega'(z, x)$ is continuous for (x) in Γ_ρ and (in view of (3.25))

$$|\omega'(z,x)| \leqq \sum_{k,i,k_1}|A_{i,k}(x)||\omega_{k,k_1}(z,x)| \to 0 \tag{3.26}$$

(as $(x)\to(z)$) *uniformly* with respect to the direction along which (x) approaches the fixed point (z) in D. Whence, on taking account of (3.15), it is inferred that

$$\Gamma^{-\frac{1}{2}}\frac{\partial\Gamma}{\partial\nu} = w(x,z)(2+\omega'(z,x)) \qquad ((x) \text{ on } S_\delta). \tag{3.27}$$

By virtue of (3.14)

$$\frac{\partial v}{\partial\nu}\delta^{m-1} = \left[\frac{2-m}{2}(w(x,z))^{m-1} + \frac{1}{\rho^{2m-4}}\left(\frac{m-2}{2}\right)\Gamma^{\frac{1}{2}m-\frac{1}{2}}\delta^{m-1}\right]\Gamma^{-\frac{1}{2}}\frac{\partial\Gamma}{\partial\nu}.$$

Since $m \geqq 3$ we have

$$\frac{\partial v}{\partial\nu}\delta^{m-1} = \left[\frac{2-m}{2}w^{m-1}(x,z) + \delta^{m-1}w_1(x,z)\right]\Gamma^{-\frac{1}{2}}\frac{\partial\Gamma}{\partial\nu}, \tag{3.28}$$

where

$$|w_1(x,z)| = \left|\frac{1}{\rho^{2m-4}}\left(\frac{m-2}{2}\right)(\sqrt{\Gamma})^{m-3}\right| \leqq w_1(z) < \infty, \tag{3.29}$$

$w_1(z)$ being independent of (x). Finally, substitution of (3.27) in (3.28) will yield

$$\frac{\partial v}{\partial\nu}\delta^{m-1} = (2-m)w^m(x,z) + R(x,z), \tag{3.30}$$

with

$$R(x,z) = 2\delta^{m-1}w_1(x,z)w(x,z) + \omega'(z,x)\left[\frac{2-m}{2}w^m(x,z) + \delta^{m-1}w_1(x,z)w(x,z)\right].$$

In view of (3.29), (3.18) and of (3.26) it is observed that $R(x, z)$ *is continuous in* (x), for (x) in Γ_ρ; moreover,

$$R(x,z) \to 0 \qquad (\text{as } (x)\to(z)),$$

uniformly with respect to the direction along which $(x)\to(z)$.

On taking account of continuity of u, in consequence of (3.30) we have

$$u(x)\frac{\partial v}{\partial \nu}\delta^{m-1} = (u(z) - u(z, x))[(2-m)w^m(x, z) + R(x, z)] \tag{3.31}$$

where $u(z, x)$ is continuous in (z), (x) ((z), (x) in D) and $\to 0$ uniformly (for (z) fixed in D), as $(x) \to (z)$.[10] Relation (3.31), together with (3.19b) (satisfied uniformly in the previously indicated sense), leads to the conclusion that for the integral (3.20b) one has

$$\lim_{\delta \to 0} Q_2(\rho, \delta) = \int^{(m-1)} \frac{u(z)(2-m)}{(\sqrt{H(z, \pi)})^m}\, d\Omega. \tag{3.32}$$

Thus, by (3.20), (3.32) and (3.24), *there is on hand the relation*

$$\lim_{\delta \to 0} Q(\rho, \delta) = -k(z)u(z), \tag{3.33}$$

where

$$k(z) = (m-2)\int^{(m-1)} \frac{d\Omega}{\sqrt{(H(z, \pi))^m}}, \tag{3.33a}$$

with

$$H(z, \pi) = \sum_{i,j} H_{i,j}(z)\pi_i \pi_j.$$

In (3.33a) $d\Omega$ is the differential element of surface of a hypersphere of radius unity, $(\pi) = (\pi_1, \cdots \pi_m)$ is thought of as representing the variable point (on this sphere) with respect to which the integration is performed.

At this stage it will be convenient to obtain certain inequalities for $k(z)$. By (3.33a) and (3.17)

$$k(z) \leqq (m-2)h^{-\frac{1}{2}m}(z)\int^{(m-1)} d\Omega = (m-2)h^{-\frac{1}{2}m}(z)S_m. \tag{3.34}$$

On the other hand,

$$H(z, \pi) \leqq H(z) < \infty \qquad \text{(for } (\pi) \text{ on unit sphere)} \tag{3.35}$$

for (z) in D; here $H(z)$ may be taken continuous (independent of (π)). Accordingly

$$k(z) \geqq (m-2)\int^{(m-1)} H^{-\frac{1}{2}m}(z)\, d\Omega = (m-2)H^{-\frac{1}{2}m}(z)S_m > 0. \tag{3.36}$$

By virtue of (3.10a) and (3.33) it is concluded that

$$-k(z)u(z) = \int_{\Gamma_\rho}^{(m)} [v(x, z)f(x) - u(x)G(v(x, z))]\, dT_x,$$

[10] Uniformly with respect to direction of approach.

where Γ_ρ is the domain, containing (z) in the interior, bounded by the surface $\Gamma(x, z) - \rho^2 = 0$ ($\rho(> 0)$ sufficiently small). Accordingly, u satisfies the integral equation

$$u(z) - \int_\Gamma^{(m)} K(x, z)u(x)\, dT_x = f_o(z), \tag{3.37}$$

with

$$f_\rho(z) = -\int_{\Gamma_\rho}^{(m)} \frac{v(x, z)}{k(z)} f(x)\, dT_x, \tag{3.37a}$$

$$K(x, z) = \frac{1}{k(z)} G(v(x, z)). \tag{3.37b}$$

We sum the above developments as follows.

THEOREM 3.1. *Every solution u of* (1.1) *which, together with the $D_1 u$ is continuous in the domain D, satisfies the integral equation* (3.37). *Here $f_\rho(z)$ is given by* (3.37a); *the kernel $K(x, z)$ is defined by* (3.37b). *The domain Γ_ρ contains (z) and is bounded by the surface $\Gamma(x, z) - \rho^2 = 0$ [$\rho(> 0)$, sufficiently small; (z) in D]. The function $v(x, z)$ involved in* (3.37a) *and* (3.37b) *is of the form*

$$v(x, z) = \Gamma^{\frac{1}{2}(2-m)}(x, z) + \frac{1}{\rho^{2m-4}} \Gamma^{\frac{1}{2}(m-2)}(x, z) - \frac{2}{\rho^{m-2}} = v(z, x).$$

The differential operator G is the adjoint of F and is, accordingly, defined by (2.1) *(the derivatives are with respect to the x_j). Finally, $k(z)$ is the function of* (3.33a) *and satisfies the inequalities* (3.34) *and* (3.36) *for (z) in D).*

We define $v(x, z)$ and, hence, $K(x, z)$ as zero for (x) (in D) exterior Γ_ρ (i.e. for (x) such that $\Gamma(x, z) \geqq \rho^2$, when $\rho(> 0)$ possibly depends on (z)). Then $K(x, z)$ will be defined for all (x) and (z) in D. With this in view, the field of integration in (3.37), (3.37a) *may be replaced by D.*

4. Investigation of the integral equation

In consequence of the developments of section 2 it is inferred that

$$\frac{\partial \Gamma(x, z)}{\partial x_k} = \sum_{k_1} (2H_{k,k_1}(x) + o_{k,k_1}(z, x))(x_{k_1} - z_{k_1}), \tag{4.1}$$

$$\frac{\partial^2 \Gamma}{\partial x_k\, \partial x_{k_1}} = 2H_{k,k_1}(x) + o'_{k,k_1}(z, x), \tag{4.1a}$$

where the functions

$$o_{k,k_1}(z, x), \qquad o'_{k,k_1}(z, x)$$

are continuous in (x) for (x) in Γ_ρ, while

$$o_{k,k_1}(z, x), \qquad o'_{k,k_1}(z, x) \to 0 \qquad (\text{as } (x) \to (z)).$$

Now

$$\sum_{i,k} \frac{\partial^2}{\partial x_i\, \partial x_k} (A_{i,k} v) = \sum_{i,k} A_{i,k} \frac{\partial^2 v}{\partial x_i\, \partial x_k} + 2 \sum_{i,k} \frac{\partial A_{i,k}}{\partial x_i} \frac{\partial v}{\partial x_k} + v \sum_{i,k} \frac{\partial^2}{\partial x_i\, \partial x_k} A_{i,k};$$

hence by (2.1)

$$G(v) = \sum_{i,k} A_{i,k}(x) \frac{\partial^2 v}{\partial x_i \partial x_k} + \sum_k B'_k(x) \frac{\partial v}{\partial x_k} + C'(x)v, \tag{4.2}$$

where

$$B'_k(x) = 2 \sum_i \frac{\partial A_{i,k}}{\partial x_i} - B_k, \tag{4.3}$$

$$C'(x) = \sum_{i,k} \frac{\partial^2}{\partial x_i \partial x_k} A_{i,k} - \sum_i \frac{\partial B_i}{\partial x_i} + C.$$

By (3.6)

$$v = v(x, z) = T(\Gamma(x, z)), \qquad T(\zeta) = \zeta^{\frac{1}{2}(2-m)} + \frac{1}{\rho^{2m-4}} \zeta^{\frac{1}{2}(m-2)} - \frac{2}{\rho^{m-2}}.$$

Thus

$$\frac{\partial v}{\partial x_k} = T^{(1)}(\Gamma) \frac{\partial \Gamma}{\partial x_k}, \qquad \frac{\partial^2 v}{\partial x_i \partial x_k} = T^{(2)}(\Gamma) \frac{\partial \Gamma}{\partial x_i} \frac{\partial \Gamma}{\partial x_k} + T^{(1)}(\Gamma) \frac{\partial^2 \Gamma}{\partial x_i \partial x_k}$$

and, by virtue of (4.2),

$$G(v) = T^{(1)}(\Gamma) \left\{ \sum_{i,k} A_{i,k}(x) \frac{\partial^2 \Gamma}{\partial x_i \partial x_k} + \sum_k B'_k(x) \frac{\partial \Gamma}{\partial x_k} \right\} + C'(x)T(\Gamma) + T^{(2)}(\Gamma) \sum_{i,k} A_{i,k}(x) \frac{\partial \Gamma}{\partial x_i} \frac{\partial \Gamma}{\partial x_k}.$$

On making use of (4.2) and of the differential equation (2.8), we accordingly obtain

$$G(v) = \tau(x, z)T^{(1)}(\Gamma) + C^1(x)T(\Gamma) + 4\Gamma T^{(2)}(\Gamma), \tag{4.4}$$

where

$$\tau(x, z) = \sum_{i,k} A_{i,k}(x) \frac{\partial^2 \Gamma}{\partial x_i \partial x_k} + \sum_k B'_k(x) \frac{\partial \Gamma}{\partial x_k} = G(\Gamma) - C'(x)\Gamma; \tag{4.4a}$$

here

$$T^{(1)}(\Gamma) = \frac{m-2}{2} \left[\frac{1}{\rho^{2m-4}} \Gamma^{\frac{1}{2}(m-4)} - \Gamma^{-\frac{1}{2}m} \right],$$

$$T^{(2)}(\Gamma) = \frac{m-2}{2} \left[\frac{1}{\rho^{2m-4}} \left(\frac{m-4}{2} \right) \Gamma^{\frac{1}{2}(m-6)} + \frac{m}{2} \Gamma^{\frac{1}{2}(-m-2)} \right]. \tag{4.4b}$$

From (4.4a) by virtue of (4.1) and (4.12) we deduce

$$\tau(x, z) = 2m + \tau_1(x, z), \tag{4.5}$$

$$\tau_1(x, z) = \sum_{i,k} A_{i,k}(x) o'_{i,k}(z, x) + \sum_{k,k_1} B'_k(x)(2H_{k,k_1}(x) + o_{k,k_1}(z, x))(x_{k_1} - z_{k_1}).$$

It is observed that $\tau(x, z)$ *is independent of* $C(x)$; moreover, $\tau(x, z)$ is continuous in (x) for (x) in Γ_ρ. Also

$$\tau_1(x, z) \to 0 \qquad \text{(as } (x) \to (z)\text{).} \tag{4.5a}$$

It is possible to obtain a more detailed expression for $\tau_1(x, z)$, namely

$$\tau_1(x, z) = \sum_j (\tau_{1,j}(z) + \bar{o}_j(z, x))(x_j - z_j); \tag{4.6}$$

here the $\tau_{i,j}(z)$ *are polynomials in the*

$$H_{i,k}(z),\ A_{i,k}(z),\ D_1 A_{i,k}(z),\ B_i(z);$$

moreover, the $\bar{o}_j(z, x)$ are continuous in (x) for (x) in Γ_ρ, while

$$\bar{o}_j(z, x) \to 0 \qquad \text{(as } (x) \to (z)\text{).} \tag{4.6a}$$

The precise form of the $\tau_{1,j}(z)$ in (4.6) is

$$\tau_{1,j}(z) = \lambda_j^0(z) + \sum_{i,k} 2H_{i,k}(z) \frac{\partial A_{i,k}(z)}{\partial z_j}, \tag{4.6b}$$

where

$$\lambda_j^0(z) = \sum_{i,\nu} A_{i,\nu}(z) N_{i,\nu}^j(z) + 2 \sum_i B_i'(z) H_{i,j}(z), \tag{4.6c}$$

with

$$N_{i,\nu}^j(z) = T_{\nu,j}^i + T_{i,\nu}^j + T_{\nu,i}^j + T_{j,\nu}^i + T_{i,j}^\nu + T_{j,i}^\nu, \tag{4.6d}$$

$$T_{k,\mu}^i = -\sum_{\tau,\gamma} H_{\tau,\mu} H_{\gamma,i} \frac{\partial A_{\gamma,\tau}}{\partial z_k} + \tfrac{1}{2} \sum_{j,\tau} H_{j,k} H_{\tau,\mu} \frac{\partial A_{j,\tau}}{\partial z_i}. \tag{4.6e}$$

By (4.4), (4.5) and (4.4b)

$$G(v) = (2m + \tau_1(x, z)) \frac{m-2}{2} \left[\frac{1}{\rho^{2m-4}} \Gamma^{\frac{1}{2}(m-4)} - \Gamma^{-\frac{1}{2}m} \right]$$

$$+ C'(x) \left[\Gamma^{\frac{1}{2}(2-m)} + \frac{1}{\rho^{2m-4}} \Gamma^{\frac{1}{2}(m-2)} - \frac{2}{\rho^{m-2}} \right] + (m-2) \left[\frac{m-4}{\rho^{2m-4}} \Gamma^{\frac{1}{2}(m-4)} + m\Gamma^{-\frac{1}{2}m} \right].$$

Thus

$$\Gamma^{\frac{1}{2}m} G(v) = -\left(\frac{m-2}{2} \right) \tau_1(x, z) + \epsilon_1(x, z) + C'(x)(1 + \epsilon_2(x, z))\Gamma(x, z), \tag{4.7}$$

where

$$\epsilon_1(x, z) = (2(m-2) + \tfrac{1}{2}\tau_1(x, z)) \frac{m-2}{\rho^{2m-4}} \Gamma^{m-2},$$

$$\epsilon_2(x, z) = \frac{1}{\rho^{2m-4}} \Gamma^{m-2} - \frac{2}{\rho^{m-2}} \Gamma^{\frac{1}{2}m-1} \qquad (\tau_1(x, z) \text{ from (4.6)–(4.6a)}). \tag{4.7a}$$

THEOREM 4.1. *The kernel of the integral equation* (3.37) *is of the form*

$$K(x, z) = \frac{1}{k(z)} G(v),$$

where $k(z)$ is given by (3.33a) (*cf.* (3.34)–(3.36)) *and $G(v)$ is of the form* (4.7), (4.7a); *moreover, $C(x)$ enters in $C'(x)$* (*cf.* (4.3)) *only.*

In view of the preceding developments and in particular in consequence of (4.6b) – (4.6e) it is inferred that, when the $A_{i,j}$ are constants, while the $B_i \equiv 0$, we have

$$\tau_{1,j}(z) \equiv 0 \qquad (j = 1, \cdots m)$$

and

$$\tau_1(x, z) = \sum_j \bar{o}_j(z, x)(x_j - z_j) \qquad \text{(cf. (4.6a))};$$

a further examination will show that, in this case, $\tau_1(x, z)$ is of the order of magnitude of r^2 $(= (x_1 - z_1)^2 + \cdots + (x_m - z_m)^2)$.

We recall that by (3.15) and (3.18)

$$\frac{r}{\sqrt{\Gamma}} \leqq \frac{1}{h_0(z)} \qquad (0 \leqq r \leqq \delta_0). \tag{4.8}$$

Now, in consequence of (4.7) and (4.7a)

$$| \Gamma^{\frac{1}{2}m} G(v) | \leqq r k_1(z, \delta_0) \qquad (r \leqq \delta_0),$$

where $k_1(z, \delta_0)$ is independent of (x). Hence from (4.8) it is deduced that

$$| G(v) | \leqq k(z, \delta_0) r^{-m+1} \qquad (r \leqq \delta_0), \tag{4.9}$$

where

$$k(z, \delta_0) = \frac{k_1(z, \delta_0)}{h_0^m(z)}$$

is independent of (x).

Choosing for the element of volume, dT_x, the expression of (3.7a), in view of (4.9) it is concluded that the integral

$$\int^{(m)} G^2(v)\, dT_x,$$

extended over a neighborhood of $(x) = (z)$, will not necessarily exist. That is, the integral

$$\int^{(m)} K^2(x, z)\, dT_x$$

may be divergent.

We shall consider now the effect of iterating the equation (3.37). It is inferred that the result of a ν-fold iteration is

$$u(z) - \int_D^{(m)} K_\nu(x, z) u(x)\, dT_x = f_{\rho,\nu}(z) \qquad (\nu = 1, 2, \cdots), \tag{4.10}$$

where

$$K_\nu(x, z) = \int^{(m)} K(x, x_1) K_{\nu-1}(x_1, z)\, dT_{x_1} \tag{4.11}$$

$$(\nu = 1, 2, \cdots ; K_0(x, z) = K(x, z))$$

and

$$f_{\rho,\nu}(z) = f_{\rho,\nu-1}(z) + \int^{(m)} K_{\nu-1}(x_1, z) f_\rho(x_1)\, dT_{x_1} \tag{4.11a}$$

$$(\nu = 1, 2, \cdots ; f_{\rho,0}(z) = f_\rho(z)).$$

In view of a known theorem[11] it is concluded that, if $K(x, z)$ is of the form

$$K(x, z) = \frac{H(x, z)}{r^\alpha} \quad (r^2 = (x_1 - z_1)^2 + \cdots + (x_m - z_m)^2; 0 < \alpha < m), \tag{4.12}$$

where $H(x, z)$ is bounded (integrable), then $K_\nu(x, z)$ is bounded (for (x), (z) in Γ_ρ) for ν such that $\nu + 1$ is the least integer for which

$$\nu + 1 > \frac{1}{1 - \frac{\alpha}{m}};$$

that is, for ν equal to the least integer for which

$$\nu > \frac{\alpha}{m - \alpha}.$$

Now, in the actual case under consideration *it is observed that,* in view of (4.9), *one has*

$$|K(x, z)| \leqq \frac{1}{k(z)} |G(v)| \leqq \frac{1}{k(z)} k(z, \delta_0) r^{-m+1}; \tag{4.13}$$

thus, (4.12) *holds with* $\alpha = m - 1$. Accordingly, *the function* $K_\nu(x, z)$ *of* (4.11) *is bounded* (and can be shown to be integrable in (x) over D) *for* $\nu = m$ *and for* (x) *in* D. The upper bound of $|K_\nu(x, z)|$ will in general depend on (z).

The integrals defining the $f_{\rho,\nu}(z)$ (4.11a) will certainly exist since $f(x)$ is bounded in any closed subset of D and is integrable over Γ_ρ. This fact is a consequence of the following considerations. By (3.6) and since

$$\Gamma(x, z) \leqq \rho^2 \qquad \text{(for } (x) \text{ in } \Gamma_\rho)$$

one has

$$v(x, z) = \Gamma^{\frac{1}{2}(2-m)}(x, z) v_0(x, z),$$

where

$$|v_0(x, z)| \leqq 4 \qquad ((x) \text{ in } \Gamma_\rho).$$

Furthermore

$$v(x, z) = w(x, z)^{m-2} v_0(x, z) r^{-m+2} \qquad \text{(cf. (3.15))}$$

and, by virtue of (3.18),

$$|v(x, z)| \leqq h'(z) r^{-m+2} \qquad (h'(z) = 4h_0^{-m+2}(z)) \tag{4.14}$$

---

[11] See, for instance, V. Volterra and J. Pérès, *Théorie genérale dés fonctionnelles . . .* [Paris, 1936; p. 275].

for (x) in Γ_ρ. On taking

$$dT_x = r^{m-1}\,dr\,d\Omega$$

and on using (4.14), we obtain for the integrand of (3.37a) the inequality

$$\left|\frac{v(x,z)}{k(z)}f(x)\,dT_x\right| \leqq \frac{h'(z)}{k(z)}\,|f(x)|\,r\,dr\,d\Omega.$$

Whence the integral (3.37a) exists. In particular, if

$$|f(x)| \leqq M \qquad \text{(in } \Gamma_\rho\text{)} \tag{4.15}$$

one has

$$|f_\rho(z)| \leqq \frac{h'(z)}{2k(z)}\,ML^2S_m \leqq f_0^*(z) \quad (L = \text{diameter of } \Gamma_\rho). \tag{4.16}$$

In view of (3.36), (4.14) we may take

$$f_0^*(z) = \frac{2ML^2}{m-2}\,H^{\frac{1}{2}m}(z)h_0^{-m+2}(z) \qquad \text{(cf. (3.35))}. \tag{4.17}$$

Here $h_0^2(z) = h(z) - \delta_0\lambda(2)$, where $h(z)$ is the minimum of $H(z,\pi)$ for (π) on the unit hypersphere; $\lambda(z)$ (from (3.16d)) may be taken as $\sigma\lambda_0(z)$ (suitable positive constant σ), with $\lambda_0(z)$ from (3.16c). In this connection, $\rho(>0)$ *is to be taken sufficiently small so that, for* (x) *in* Γ_ρ, *one has* $r \leqq \delta_0 < h(z)/\lambda(z)$.

Let

$$[z]$$

denote generically a positive function of (z) uniformly bounded in every closed subset of D. On taking account of (4.11a) (with $\nu = 1$) and of the fact that, in consequence of (4.13)

$$|K(x_1, z)| \leqq [z]r_1^{-m+1}$$

(r_1 = distance between (z) and (x_1)) we have (on taking $dT_{x_1} = r_1^{m-1}\,dr_1\,d\Omega_1$)

$$|K(x_1, z)f_\rho(x_1)\,dT_{x_1}| \leqq [z][x_1]\,dr_1\,d\Omega_1,$$

it is concluded that $|f_{\rho_1}(z)| = [z]$. In view of (4.11a) (with $\nu = 2$) and since

$$|K_1(x_1, z)f_\rho(x_1)\,dT_{x_1}| \leqq [z][x_1]r_1^{-m_1}\,dT_{x_1} \qquad (-m_1 \geqq -m+1),$$

so that

$$|K_1(x_1, z)f_\rho(x_1)\,dT_{x_1}| \leqq [z][x_1]r_1^{-m_1+m-1}\,dr_1\,d\Omega_1,$$

we infer that, inasmuch as $-m_1 + m - 1 \geqq 0$, necessarily $|f_{\rho,2}(z)| = [z]$. Thus, on using the fact that

$$|K_\nu(x, z)| \leqq [z]r^{-m_\nu},$$

where

$$-m+1 \leqq -m_1 \leqq -m_2 \leqq \cdots,$$

from (4.11a) we deduce, step by step, that

$$|f_{\rho,\nu}(z)| = [z]$$

for $\nu = 1, \cdots m$. In the above we use the fact that for every (z) in D, there is a closed subset of D (containing (z)) so that $K_\nu(x, z)$ is zero for (x) exterior this subset.

THEOREM 4.2. *Every solution u of the partial differential equation* (1.1), *which together with the $D_1 u$ is continuous in the domain D, satisfies the integral equation*

$$u(z) - \int_{\Gamma_{\rho'}}^{(m)} K_m(x, z)u(x)\, dT_x = f_{\rho,m}(z) \qquad (\rho' > \rho), \tag{4.18}$$

where $K_m(x, z)$ and $f_{\rho,m}$ are defined with the aid of the relations (4.11), (4.11a). *This equation has the advantage over the original equation* (3.37) *in the fact that the kernel $K_m(x, z)$ satisfies the inequality*

$$|K_m(x, z)| \leqq k^*(z) < \infty \qquad (k^*(z) \text{ independent of } (x))$$

for (x) in $\Gamma_{\rho'}$, and that $K_m(x, z)$ is integrable, in (x), over $\Gamma_{\rho'}$.

NOTE. By taking $\rho = \rho(z)$ (> 0) suitably small we arrange to have $\rho' = \rho'(z) > \rho(z)$, above, so small that the domain $\Gamma_{\rho'}$ (i.e. the set of points (x) such that $\Gamma(x, z) \leqq (\rho')^2$) lies in D. *We have $K_m(x, z)$ zero for (x) (in D) exterior $\Gamma_{\rho'}$; in* (4.18) *we may then replace the field of integration by D.* The occurrence of ρ' takes place in consequence of the following considerations. By (4.11) (with $\nu = 1$)

$$K_1(x, z) = \int_D^{(m)} K(x, x_1)K(x_1, z)\, dT_{x_1} = \int_{\Gamma_\rho}^{(m)} K(x, x_1)K(x_1, z)\, dT_{x_1} \tag{4.19}$$

since $K(x_1, z) = 0$ for (x_1) exterior Γ_ρ. Now $K(x, x_1) = 0$ for (x) (in D) such that $\Gamma(x, x_1) \geqq \rho(x_1) > 0$. There exists a number $\rho_1 = \rho_1(z) > \rho(z)$, independent of (x_1) ((x_1) in Γ_ρ) so that $K(x, x_1) = 0$ for (x) (in D) exterior Γ_{ρ_1}, for all (x_1) in Γ_ρ. Hence $K_1(x, z) = 0$ for (x) exterior Γ_{ρ_1}; $\rho_1(z)$ can be made as small as desired by suitable choice of $\rho = \rho(z)$. By (4.11) (with $\nu = 2$)

$$K_2(x, z) = \int_D^{(m)} K(x, x_1)K_1(x_1, z)\, dT_{x_1} = \int_{\Gamma_{\rho_1}}^{(m)} K(x, x_1)K_1(x_1, z)\, dT_{x_1}. \tag{4.19a}$$

Repeating the above reasoning, with ρ_1 in place of ρ we find that $K_2(x, z) = 0$ for (x) exterior Γ_{ρ_2}, where $\rho_2 = \rho_2(z)$ $(> \rho_1(z))$. Proceeding step by step in such a manner, we obtain the stated property of $K_m(x, z)$, with $\rho' = \rho(z)$.

In the sequel the prime over ρ will be deleted.

If the D_n are domains, in D, such that $\bar{D}_n \subset D_{n+1}$ and $D_n \to D$ (as $n \to +\infty$), with the aid of the above we obtain the following result:

$$u(z) - \int_{D_{n'}}^{(m)} K_m(x, z)u(x)\, dT_x = f_{\rho,m}(z) \tag{4.20}$$

for all (z) *in* D_n ; *here* n' *is some integer* $> n$ such that $D_{n'}$ contains every point (x) for which $\Gamma(x, z) \leqq \rho^2(z)$, when (z) is any point in $\bar{D}_n$; *with* $\rho(z)$ *suitably small such* n' *certainly exists and, in fact, may be taken as* $n + 1$. *Clearly* $K_m(x, z) = 0$ *for* (z) *in* D_n *and* (x) *in* $D - \bar{D}_n$.

We have $k^*(z) \leqq b_n < \infty$ *(for* (z) *in* $\bar{D}_n$*), where* $k^*(z)$ *is from* (4.18a). *Thus*

$$| K_m(x, z) | \leqq b_n < \infty \qquad \textit{(constant } b_n\textit{)} \tag{4.20a}$$

for all (x) *in* D *and for all* (z) *in* D_n .

5. Characteristic values and functions

In place of (1.1) we shall now consider the partial differential equation

$$F(u) + \lambda u = f \qquad (F \text{ from (1.1)}; \text{ continuous } f \subset L_2 \text{ in } D), \tag{5.1}$$

where λ is a parameter. This equation is obtained by replacing C by $C + \lambda$. Any equation of the form

$$F(u) + \lambda q(x) u = f$$

is reduced to the form (5.1), provided that, for (x) in D, the functions

$$A_{i,k}\sigma, \qquad B_i\sigma, \qquad C\sigma, \qquad f\sigma \qquad \left(\text{with } \sigma = \frac{1}{q(x)}\right)$$

satisfy conditions as previously imposed on the functions

$$A_{i,k}\,,\ B_i\,,\ C, f,$$

respectively.

The introduction of a parameter, as above, is in agreement with the situation in the Schrödinger wave theory.

On taking account of (3.37b), $v(x, z)$ being of the form (3.6), and of (2.1), replacing C by $C + \lambda$ we obtain

$$K(x, z) = K^0(x, z) + \frac{\lambda}{k(z)}\, v(x, z), \tag{5.2}$$

where $K^0(x, z)$ stands for the function which in section 4 has been designated as $K(x, z)$.

The function $K(x, z)$, defined in (5.2), is the kernel of the integral equation (3.37), corresponding to the partial differential equation (5.1).

Application of (4.11), for $\nu = 1, \cdots m$, will yield

$$K_m(x, z) = K_m^0(x, z) + \sum_{j=1}^{m+1} \lambda^j A_{m,j}(x, z). \tag{5.3}$$

Here $K_m^0(x, z)$ *is the function designated in section* 4 *as* $K_m(x, z)$. *The* $A_{m,j}(x, z)$ *are functions of the same type as* $K_m(x, z)$ *of section* 4 *and accordingly satisfy inequalities of the form* (4.20a).

In view of (337a), (4.11) and (4.11a) it is observed that $f_{\rho,m}(z)$ *is a polynomial*

in λ *of degree* m, *whose coefficients are uniformly bounded* (*in absolute value*) *and integrable over* Γ_ρ. In order to put in evidence dependence on λ we shall write

$$f_{\rho,m}(z) = f_{\rho,m}(z, \lambda).$$

The integral equation corresponding to (5.1) is

$$(5.4) \qquad u(z) - \int_{\Gamma_\rho}^{(m)} u(x)\left[K_m^0(x, z) + \sum_{j=1}^{m+1} \lambda^j A_{m,j}(x, z)\right] dT_x = f_{\rho,m}(z, \lambda).$$

Before proceeding further we shall note that in consequence of a recent work of G. Giraud,[12] in the sequel referred to as (G), the following may be asserted with regard to the problem

$$(5.5) \qquad F(u) + \lambda u = f \qquad ((x) \text{ in } D_n),$$

$$(5.5a) \qquad \frac{\partial u}{\partial \nu} + a_n(x)u = p_n(x) \qquad ((x) \text{ on } S_n),$$

where D_n is a domain, for which $\bar{D} \subset D$, and S_n is the limiting surface of D_n; the functions $a_n(x)$, $p_n(x)$ are defined and continuous on S_n. It is assumed that D_n and S_n are *"regular"* in the sense that S_n can be covered by a finite number of domains, in each of which one of the Cartesian coördinates of a point (variable in S_n) is expressible with the aid of $m - 1$ other coördinates through a function whose partial derivatives of the first order belong to the class Lip. h (with $h \leqq 1$).[13]

In (G) the problem (5.5), (5.5a) (and, in fact, a more general problem) is reduced to integral equations to which the Fredholm theory applies. The final conclusion of (G), needed in the sequel, is that "either the only solution of the homogeneous problem is zero, and then the non homogeneous problem is compatible for all f (continuous in $D_n + S_n$) and $p_n(x)$; or the homogeneous problem has solutions not identically zero, deducible from p (p finite) linearly independent solutions, and then the non homogeneous problem is not compatible, unless certain p necessary and sufficient conditions are satisfied." The latter conditions are supplied by the Fredholm theory.

We shall take $p_n(x) \equiv 0$. Associated with (5.5), (5.5a) there is a set of *characteristic values*,

$$(5.6) \qquad \lambda_{n,i} \qquad (i = 1, 2, \cdots),$$

the sequence $\lambda_{n,1}, \lambda_{n,2}, \cdots$ possessing no finite limiting values. When $\lambda \neq \lambda_{n,i}$,

---

[12] G. Giraud, *Nouvelle méthode pour traiter certains problèmes relatifs aux équations du type elliptique* [Journal de Mathématiques, Tome Dix-huitième (1939), 111–143].

[13] This is the type of surfaces considered in (*G*).

then the non homogeneous problem (5.5), (5.5a) has a solution, $u = u_n$, continuous in $D_n + S_n$. For $\lambda = \lambda_{n,i}$ the homogeneous problem

$$\begin{aligned} &F(u) + \lambda_{n,i}u = 0 && \text{(in } D_n\text{)}, \\ &\frac{\partial u}{\partial \nu} + a_n(x)u = 0 && \text{(on } S_n\text{)} \end{aligned} \tag{5.7}$$

has at least one solution, $u = u_{n,i}$, distinct from zero, continuous in $D_n + S_n$.

We assume that D is such that there exists a sequence of domains

$$D_n \qquad (n = 1, 2, \cdots ; \bar{D}_n \subset D),$$

whose limiting surfaces are S_n $(n = 1, 2, \cdots)$, respectively, and which are "regular" in the sense specified above, while

$$\bar{D}_n \subset D_{n+1}, \qquad \lim_n D_n = D.^{14}$$

Forthwith the D_n and S_n will have the meaning just indicated.

In connection with (5.1) one may consider the "adjoint" non homogeneous equation

$$G(v) + \lambda v = g \qquad \text{(continuous } g \subset L_2 \text{ in } D\text{)}.$$

With respect to this equation results will hold precisely analogous to those stated from (5.1) to (5.4) for the equation (5.1).

Furthermore, applying the work of Giraud to the problem

$$\begin{aligned} &G(v) + \rho v = g && \text{(in } D_n\text{)}, \\ &\frac{\partial v}{\partial \nu} + b_n(x)v = q_n(x) && \text{(on } S_n\text{)}, \end{aligned} \tag{5.8}$$

where $b_n(x)$, $q_n(x)$ are continuous on S_n, we have a situation similar to that described in connection with (5.5), (5.5a). In particular, the problem (5.8) has a set of *characteristic values*

$$\rho_{n,j} \qquad (j = 1, 2, \cdots), \tag{5.9}$$

the sequence $\rho_{n,1}, \rho_{n,2}, \cdots$ having no finite limiting points, such that the homogeneous problem

$$\begin{aligned} &G(v) + \rho_{n,j}v = 0 && \text{(in } D_n\text{)}, \\ &\frac{\partial v}{\partial \nu} + b_n(x)v = 0 && \text{(on } S_n\text{)} \end{aligned} \tag{5.10}$$

has a solution, $v = v_{n,j}$, distinct from zero, continuous in $D_n + S_n$.

---

[14] Here $\lim D_n = D_1 + D_2 + \cdots$.

Under suitable general conditions on u and v, in view of the fundamental formula (2.7) we have

$$(5.11)\qquad \int_{D_n}^{(m)} [vF(u) - uG(v)]\,dT = -\int_{S_n}^{(m-1)} \left[v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu} + L_n uv\right] dS,$$

where (cf. 2.7a)) L_n is defined on S_n by

$$(5.12)\qquad L_n = \sum_i \pi_i(n) B_i - \sum_{i,k} \pi_i(n)\frac{\partial A_{i,k}}{\partial x_k};$$

here the $\pi_i(n)$ are direction cosines of normal to the surface S_n. In view of the previously made suppositions with respect to the B_i, $A_{i,k}$ and the surface S_n, the function L_n is continuous on S_m.

It is observed that, if v satisfies the second relation (5.10) and u satisfies the last equation (5.7), one has

$$v\frac{\partial u}{\partial \nu} - u\frac{\partial v}{\partial \nu} + L_n uv \equiv (b_n - a_n + L_n)uv \qquad \text{(on } S_n\text{)}.$$

Hence in order to make the above expression vanish on S_n we shall take

$$(5.13)\qquad b_n = a_n - L_n\,.$$

THEOREM 5.1. 1°. *The set of characteristic values* $\{\lambda_{n,i}\}$ [(5.6)] *is identical with the set of characteristic values* $\{\rho_{n,j}\}$ [(5.9)].[15] *We write* $\lambda_{n,i} = \rho_{n,i}$.

2°. *When F is self adjoint the characteristic values are all real and the characteristic functions may be taken real.*

3°. *There are on hand orthogonality relations*:

$$(5.14)\qquad \int_{D_n}^{(m)} u_{n,i}(x)v_{n,j}(x)\,dT = 0 \qquad (\textit{for } \lambda_{n,i} \neq \lambda_{n,j});$$

$$(5.14a)\qquad \int_{D_n}^{(m)} u_{n,i}(x)u_{n,j}(x)\,dT = 0 \qquad (\textit{for } \lambda_{n,i} \neq \lambda_{n,j},\ \textit{when } F \textit{ is self adjoint}).$$

To prove (1°) consider a characteristic value $\lambda_{n,i}$ for (5.7); thus,

$$(5.15)\qquad \begin{aligned} F(u_{n,i}) + \lambda_{n,i}u_{n,i} &= 0 && \text{(in } D_n\text{)},\\ \frac{\partial u_{n,i}}{\partial \nu} + a_n(x)u_{n,i} &= 0 && \text{(on } S_n\text{)}, \end{aligned}$$

where $u_{n,i}$ is a characteristic function. Suppose, if possible, that $\lambda_{n,i}$ is not a characteristic value for (5.10) (under (5.13)).[16] Then every non homogeneous problem

$$\begin{aligned} G(v) + \lambda_{n,i}v &= g && \text{(in } D_n\text{)},\\ \frac{\partial v}{\partial \nu} + b_n v &= 0 && \text{(on } S_n\text{)}, \end{aligned}$$

[15] Nothing is said about the multiplicities of these values.

[16] We take $n_1 < n_2 < \cdots$; $n_s \to \infty$ with s.

where $\lambda_{n,i}$ is fixed but g is any function continuous in $D_n + S_n$, will have a solution v. Now substitute this, supposedly existent, function v, as well as $u = u_{n,i}$, in (5.11). One obtains

$$0 = \int_{D_n}^{(m)} [vF(u_{n,i}) - u_{n,i}G(v)]\,dT = -\int_{D_n}^{(m)} u_{n,i}g\,dT,$$

which implies that $u_{n,i} = 0$ in D_n, contrary to the assertion made subsequent to (5.7). Hence $\lambda_{n,i}$ is a characteristic value for the "associated" homogeneous boundary value problem. The converse is proved by a similar method.

Substitute in (5.11) $u = u_{n,i}$, $v = v_{n,j}$. In view of the remark preceding (5.13) the integrand of the second member in (5.11) is zero on S_n; accordingly

$$0 = \int_{D_n}^{(m)} [v_{n,j}F(u_{n,i}) - u_{n,i}G(v_{n,j})]\,dT = (\lambda_{n,j} - \lambda_{n,i}) \int_{D_n}^{(m)} u_{n,i}v_{n,j}\,dT. \tag{5.16}$$

This implies (5.14).

In view of (4.2) and (4.3) the case when F is self adjoint, that is when $F = G$, is on hand if and only if

$$B_k(x) = \sum_i \frac{\partial A_{i,k}(x)}{\partial x_i} \qquad \text{(in } D\text{)}. \tag{5.17}$$

Thus in the self adjoint case it follows by (5.12) that

$$L_n = 0.$$

Under (5.13), the homogeneous boundary value problems (5.7) and (5.10) become identical. In consequence of these considerations and of (5.14) it is inferred that (5.14a) holds as stated.

With F self adjoint suppose, if possible, that there is a non real characteristic value $\lambda_{n,i}$. On taking conjugates of the members in (5.15) we deduce that $\bar{\lambda}_{n,i}$ will be another characteristic value, while $\bar{u}_{n,i}$ will be a characteristic function corresponding to $\bar{\lambda}_{n,i}$. By virtue of (5.16), where we put

$$\lambda_{n,j} = \bar{\lambda}_{n,i}, \qquad v_{n,j} = \bar{u}_{n,i},$$

it is concluded that

$$(\bar{\lambda}_{n,i} - \lambda_{n,i}) \int_{D_n}^{(m)} |u_{n,i}|^2\,dT = 0$$

and that, accordingly, $\bar{\lambda}_{n,i} - \lambda_{n,i} = 0$. This is impossible with $\lambda_{n,i}$ non real. Thus part 2° of the Theorem has been established.

In the self adjoint case the $u_{n,i}$ $(i = 1, 2, \cdots)$ will be always arranged as a set orthonormal in D_n; we define

$$u_{n,i}(x) = 0 \qquad \text{(in } D - (D_n + S_n)\text{)};$$

the $u_{n,i}(x)$ will form a set orthonormal in D.

When F is not self adjoint orthonormalization of the $u_{n,i}$ $(i = 1, 2, \cdots)$ will yield a sequence whose members in general will not be characteristic func-

tion. This apparently involves the main reason for the difficulty of the treatment of equations which are not necessarily self adjoint—at least in certain questions relating to such equations.

The set of numbers

$$\lambda_{n,i} \qquad (n, i = 1, 2, \cdots)$$

is at most denumerably infinite. Hence there exists a real number λ_0 distinct from all of the above numbers. One may write

$$F(u) + \lambda u \equiv F^*(u) + \lambda^* u, \qquad G(u) + \lambda u \equiv G^*(u) + \lambda^* u,$$

where

$$F^* = F + \lambda_0, \qquad G^* = G + \lambda_0, \qquad \lambda^* = \lambda - \lambda_0.$$

This amounts to augmenting the coefficients of u in F and G by λ_0. In view of these considerations, augmenting, if necessary, C (and, hence $C' = C'' + C$, where C'' is independent of C) by a suitable λ_0 and retaining the original notation, *we may consider that the* $\lambda_{n,i}$ $(n, i = 1, 2, \cdots)$ *are all distinct from zero*; clearly, this entails no loss of generality.

In the sequel we confine ourselves to the case when F *is self adjoint and when, accordingly,* (5.17) *holds.*

6. The non homogeneous self adjoint problem

Let us first consider equation (5.1) for a non real value of λ. Then, in view of the reality of the characteristic values (cf. 2° of Theorem 5.1) of the problem (5.7), it can be asserted that the problem (with continuous f possibly complex valued, $\subset L_2$ in D)

$$\begin{aligned} &F(u) + \lambda u = f && \text{(in } D_n\text{)}, \\ &\frac{\partial u}{\partial \nu} + a_n(x)u = 0 && \text{(on } S_n\text{)} \end{aligned} \tag{6.1}$$

has a continuous solution u_n in $D_n + S_n$. The function $\bar{u}_n$ will satisfy

$$\begin{aligned} &F(\bar{u}_n) + \bar{\lambda}\bar{u}_n = \bar{f} && \text{(in } D_n\text{)}, \\ &\frac{\partial \bar{u}_n}{\partial \nu} + a_n(x)\bar{u}_n = 0 && \text{(on } S_n\text{)}. \end{aligned}$$

Accordingly application of (5.11) to

$$u = u_n, \qquad v = \bar{u}_n$$

will yield

$$0 = \int_{D_n}^{(m)} [\bar{u}_n F(u_n) - u_n F(\bar{u}_n)]\, dT$$

and, hence,

$$\int_{D_n}^{(m)} |u_n|^2\, dT = \frac{1}{(\lambda - \bar{\lambda})} \int_{D_n}^{(m)} (\bar{u}_n f - u_n \bar{f})\, dT.$$

By the Schwarzian inequality

$$\int_{D_n}^{(m)} |u_n|^2 dT \leqq \frac{2}{|\lambda - \bar{\lambda}|}\left[\int_{D_n}^{(m)} |u_n|^2 dT\right]^{\frac{1}{2}}\left[\int_{D_n}^{(m)} |f|^2 dT\right]^{\frac{1}{2}}$$

and, finally,

$$\int_{D_n}^{(m)} |u_n|^2 dT \leqq \frac{4}{|\lambda - \bar{\lambda}|^2}\int_{D_n}^{(m)} |f|^2 dT \leqq \frac{4}{|\lambda - \bar{\lambda}|^2}\int_{D}^{(m)} |f|^2 dT = \alpha(\lambda) \tag{6.2}$$

($n = 1, 2, \cdots$; $\alpha(\lambda)$ independent of n and $< +\infty$). *We define*

$$u_n(x) = 0 \qquad (\text{in } D - (D_n + S_n)). \tag{6.3}$$

Then

$$\int_{D}^{(m)} |u_n|^2 dT \leqq \alpha(\lambda) \qquad (n = 1, 2, \cdots; \text{cf. (6.2)}). \tag{6.4}$$

This is analogous to a procedure in (C).

In consequence of a theorem of F. Riesz, inequalities of the form (6.4) imply that there exists an infinite subsequence

$$u_{n_i}(x) \qquad (i = 1, 2, \cdots)$$

and a function $u(x)$, for which

$$\int_{D}^{(m)} |u(x)|^2 dT \leqq \alpha(\lambda) \tag{6.5}$$

such that

$$u_{n_i}(x) \to u(x) \qquad (\text{in } D; \text{ as } n_i \to \infty)$$

in the *weak* sense.

Now, inasmuch as $u_{n_i}(x)$ is a solution of (5.1), in D_{n_i}, u_{n_i} satisfies the integral equation (5.4); thus with $r > i$

$$u_{n_r}(z) = \int_{D_{n_{i+1}}}^{(m)} u_{n_r}(x)\left[K_m^0(x, z) + \sum_{s=1}^{m+1} \lambda^s A_{m,s}(x, z)\right] dT_x + f_{\rho,m}(z, \lambda) \tag{6.6}$$

for (z) in D_{n_i} ($\rho(z)$, >0, being suitably chosen so that the kernel vanishes for (x) in $D - \bar{D}_{n_{i+1}}$, when (z) is in D_{n_i}).

We recall now a theorem of F. Riesz which, for a single variable, asserts that if

$$f_\nu(x) \subset L_2 \quad (\text{on } (a, b); \nu = 1, 2, \cdots), \qquad f_\nu(x) \to f(x) \qquad (\textbf{weakly})$$

and if

$$g(x) \subset L_2 \qquad (\text{on } (a, b))$$

then

$$\lim_{\nu} \int_a^b f_\nu(x) g(x)\, dx = \int_a^b f(x) g(x)\, dx.$$

An obvious extension of this theorem to the m-space of the point (x) enables us to assert that

$$\lim_r \int_{D_{n_i+1}}^{(m)} u_{n_r}(x)K_m(x, z)\,dT_x = \int_{D_{n_i+1}}^{(m)} u(x)K_m(x, z)\,dT_x \tag{6.7}$$

(for (z) in D_{n_i}), if

$$K_m(x, z) \subset L_2 \qquad \text{(in } (x), \text{ over } D).$$

Now the latter property certainly holds by the preceding.

Thus, given any (z) in D, we choose i so that D_{n_i} contains (z); from (6.6) and (6.7) we then infer that

$$\lim u_{n_r}(z) = u'(z) \tag{6.8}$$

exists in the ordinary sense for (z) in D; clearly

$$u'(z) = u(z) \qquad \text{(in } D) \tag{6.9}$$

and

$$u(z) - \int^{(m)} u(x)K_m(x, z)\,dT_x = f_{\rho,m}(z, \lambda).$$

In (6.9) $u(z)$ is the function for which inequality (6.5) has been asserted.

We have the following theorem.

THEOREM 6.1. *Consider the partial differential equation* (5.1), *with* $f(x)$ *continuous and* $\subset L_2$ *in* D; *let* F *be self adjoint.*

For every non real value of the parameter λ (5.1) *possesses a solution* $u(x)$ *such that*

$$\int_D^{(m)} |\,u(x)\,|^2\,dT \leqq \alpha(\lambda) \qquad (\alpha(\lambda) \textit{ from } (6.2)).$$

This theorem is analogous to a result established in (C) for the case when (1.3) holds.

We now wish to investigate the more complicated situation when λ is allowed to be real. In this there will be a certain analogy with an earlier work of Trjitzinsky.[4]

For λ distinct from the $\lambda_{n,i}$ $(i = 1, 2, \cdots)$ the problem (6.1) has a solution $u_n(x)$. On substituting $u = u_n$, $v = u_{n,i}$ in (5.11) and noting (5.15) and the fact that $L_n = 0$ we obtain

$$\int_{D_n}^{(m)} [u_{n,i}F(u_n) - u_nF(u_{n,i})]\,dT = 0$$

and

$$\int_{D_n}^{(m)} u_n(x)u_{n,i}(x)\,dT = \frac{1}{\lambda - \lambda_{n,i}} \int_{D_n}^{(m)} f(x)u_{n,i}(x)\,dT.$$

Thus, by (6.3)

$$\int_D^{(m)} u_n(x)u_{n,i}(x)\,dT = \frac{1}{\lambda - \lambda_{n,i}} \int_D^{(m)} f(x)u_{n,i}(x)\,dT. \tag{6.10}$$

It is noted that (6.10) constitutes a relation between the Fourier coefficients, with respect to the $u_{n,i}$ $(i = 1, 2, \cdots)$, of the functions $u_n(x)$ and $f(x)$.

Now *an orthonormal set* $\{u_{n,i}\}$ $(i = 1, 2, \cdots)$ *is necessarily complete.* This can be established by an extension of the familiar methods, involved in known analogous, though much simpler, boundary value problems.

On writing

$$u^{n,i} = \int_D^{(m)} u_n(y) u_{n,i}(y)\, dT, \tag{6.11}$$

we have

$$\int_D^{(m)} |\, u_n(x)\,|^2\, dT = \sum_i |\, u^{n,i}\,|^2. \tag{6.11a}$$

In view of (6.10)

$$S_n^2 = \sum_i |\, u^{n,i}\,|^2 = \sum_i \left| \frac{f^{n,i}}{\lambda - \lambda_{n,i}} \right|^2, \qquad f^{n,i} = \int_D^{(m)} f(x) u_{n,i}(x)\, dT. \tag{6.12}$$

DEFINITION 6.1. *We designate by T a set of points λ, in the complex λ-plane, such that*

$$\sum_i \left| \frac{f^{n,i}}{\lambda - \lambda_{n,i}} \right|^2 \leqq B(\lambda) < +\infty \qquad (n = n_1, n_2, \cdots; f^{n,i} \text{ from } (6.12)^{16} \tag{6.13}$$

for all λ in T, the function $B(\lambda)$ being independent of n.

Let us form the complement, with respect to the λ-plane, of the set of points

$$\lambda_{n,i} \qquad (n = n_1, n_2, \cdots; i = 1, 2, \cdots). \tag{6.14}$$

We define O as the set of interior points of this complement. Clearly O contains all the points not on the axis of reals. O will contain also some points on the axis of reals if the complement, with respect to the axis, of the set of points (6.14) *contains interior (with respect to the axis) points—the latter points will belong to O.*[17] With δ positive, let O_δ be the subset of O consisting of points at the distance $\geqq \delta$ from the frontier of O; then O_δ will be a closed two dimensional set; moreover,

$$|\, \lambda - \lambda_{n,i}\,| \geqq \delta \qquad (n = n_1, n_2, \cdots; i = 1, 2, \cdots) \tag{6.15}$$

whenever λ is in O_δ. In view of (6.15) we have

$$\sum_i \left| \frac{f^{n,i}}{\lambda - \lambda_{n,i}} \right|^2 \leqq \frac{1}{\delta^2} \sum_i |\, f^{n,i}\,|^2 \qquad (\text{in } O_\delta)$$

and, in consequence of the second relation (6.12) as well as of Bessel's inequality,

$$\sum_i \left| \frac{f^{n,i}}{\lambda - \lambda_{n,i}} \right|^2 \leqq \frac{1}{\delta^2} \int_D^{(m)} |\, f(x)\,|^2\, dT \qquad (\text{in } O_\delta).$$

---

[17] It is conceivable to have the set (6.14) everywhere dense on the axis of reals, in which case, by definition, O will consist of the λ-plane minus the axis of reals. However, the set T of Definition 6.1 may have points on the axis of reals, even under these circumstances, provided the functions f are suitably chosen.

The set O, defined in connection with (6.14), *is a set T of Definition* 6.1 *with*

$$B(\lambda) = \frac{1}{\delta^2}\int_D^{(m)} |f(x)|^2\, dT$$

for λ in O; here δ is the distance from the point representing λ to the frontier of O.

An important feature of sets T of the special form O lies in the fact that they are independent of the choice of f—in so far as $f \subset L_2$ (in D); this is not so for the general sets T. In any case, of course, T includes all the points not on the axis of reals.

By (6.12) and (6.13)

$$S_n^2 \leqq B(\lambda) \qquad (n = n_1, n_2, \cdots; \text{in } T)$$

and, in view of (6.11a),

$$\int_D^{(m)} |u_n(x)|^2\, dT \leqq B(\lambda) < +\infty \qquad (n = n_1, n_2, \cdots) \tag{6.16}$$

for λ in T. The second member, here, being independent of n, one may assert that the sequence $(n_1, n_2, \cdots)$ contains a subsequence $(k_1, k_2, \cdots)$ so that

$$u_{k_j}(x) \to u(x) \qquad (\text{as } k_j \to \infty; \text{in } D),$$

convergence being in the weak sense to a function $u(x)$ for which

$$\int_D^{(m)} |u(x)|^2\, dT \leqq B(\lambda).$$

By a reasoning employed before and on making use of the integral equation it is concluded that $u(x)$ is a solution of the non homogeneous problem (5.1) for λ in T.

We have the following Existence Theorem.

THEOREM 6.2. *Let F be self adjoint (in D). The $\lambda_{n,i}$ $(n = n_1, n_2, \cdots; i = 1, 2, \cdots)$ are the characteristic numbers of a sequence of approximating boundary value problems; together with the characteristic functions $u_{n,i}$ the $\lambda_{n,i}$ satisfy* (5.15). *In terms of the $u_{n,i}(x)$ $(u_{n1}, u_{n,2}, \cdots$ orthonormalized in $D)$ we define numbers $f^{n,i}$ by the second relations* (6.12). *Let T be a set for which* (6.13) *of Definition* 6.1 *holds.*

Then the non homogeneous problem

$$F(u) + \lambda u = f$$

will have a solution u, in D, $\subset L_2$ in D for every λ in T.

DEFINITION 6.2. *It will be said that F is of class* I *on a set O, if every solution, $\subset L_2$ in D, of the homogeneous problem*

$$F(u) + \lambda u = 0$$

is zero in D—this being so for every λ in O. In the contrary case F will be said to be of class II.

Designation of classes I, II, here, is suggested by an analogous usage in *Carleman's* theory of integral equations.

It is to be recalled that corresponding to the equation (5.1) we have an iterated integral equation (5.4); the latter may be written in the form

$$u(z) - \int_D^{(m)} u(x) \sum_{i=0}^{m+1} \lambda^i A_{m,i}(x, z)\, dT_x = f_{\rho,m}(z, x) \qquad (A_{m,0} = K_m^0). \tag{6.17}$$

On taking account of the text in connection with (4.20) we note that in (6.17) the field of integration may be taken as D_{n+1} if (z) is in $\bar{D}_n$. *The expression*

$$a(z \mid D) = \sum_{i=0}^{m+1} \left[\int_D^{(m)} A_{m,i}^2(z_1, z)\, dT_z \right]^{\frac{1}{2}} \tag{6.18}$$

defines a function for all (z) *in* D. In fact, let (z) be a fixed point in D. Then for some n the domain D_n contains (z); for all (z_1) in $D - \bar{D}_{n+1}$ we have $A_{m,i}(z_1, z) = 0$. Hence

$$\int_D^{(m)} A_{m,i}^2(z_1, z)\, dT_{z_1} = \int_{D_{n+1}}^{(m)} A_{m,i}^2(z_1, z)\, dT_{z_1}.$$

Here

$$| A_{m,i}(z_1, z) | \leqq \gamma_n < \infty \qquad \text{(all } (z) \text{ in } D_n \text{ ; all } (z_1) \text{ in } D)$$

in view of the statement with respect to (4.20a). Thus the integral above exists.

7. Spectral theory

The "spectrum" of F, with respect to the boundary value problem

$$\begin{aligned} F(u) + \lambda u &= 0 && \text{(in } D_n \text{ ; parameter } \lambda), \\ \frac{\partial u}{\partial \nu} + a_n(x) u &= 0 && \text{(on } S_n), \end{aligned} \tag{7.1}$$

we define to be the function $\theta_n(x, y \mid \lambda)$ for which

$$\begin{aligned} \theta_n(x, y \mid \lambda) &= \sum_{0 < \lambda_{n,i} < \lambda} u_{n,i}(x) u_{n,i}(y) && \text{(for } \lambda > 0), \\ \theta_n(x, y \mid \lambda) &= -\sum_{\lambda \leqq \lambda_{n,i} < 0} u_{n,i}(x) u_{n,i}(y) && \text{(for } \lambda < 0), \end{aligned} \tag{7.2}$$

while $\theta_n(x, y \mid 0) = 0$; this function is defined for (x), (y) in $D_n + S_n$ and for all real λ. Clearly $\theta_n(x, y \mid \lambda) = \theta_n(y, x \mid \lambda)$. Such definition is possible inasmuch as it had been previously arranged to have the $\lambda_{n,i}$ all distinct from zero and since the $\lambda_{n,i}$ are all real.

In consequence of the definition of $u_{n,i}$ we have

$$\theta_n(x, y \mid \lambda) = 0$$

for (x) exterior $D_n + S_n$, as well as for (y) exterior $D_n + S_n$.

From the definition of θ_n it follows that

$$\Delta_{\lambda_1}^{\lambda_2} \theta_n \equiv \theta_n(x, y \mid \lambda_2) - \theta_n(x, y \mid \lambda_1) = \sum_{\lambda_1 \leqq \lambda_{n,i} < \lambda_2} u_{n,i}(x) u_{n,i}(y)$$

when $\lambda_2 > \lambda_1$. Thus, on taking

$$\lambda_0 < \lambda_1 < \cdots < \lambda_q,$$

we obtain

$$\delta_n = \sum_{\nu=1}^{q} |\Delta_{\lambda_{\nu-1}}^{\lambda_\nu} \theta_n(x, y \mid \lambda)| \leqq \sum_{\lambda_0 \leqq \lambda_{n,i} < \lambda_q} |u_{n,i}(x) u_{n,i}(y)| \tag{7.3}$$

for (x) and (y) in D.

In view of the first relation (5.15) and of the previously established connection with integral equations we have

$$u_{n,i}(z) = \sum_{j=0}^{m+1} \lambda_{n,i}^{j} \int_{D_n}^{(m)} u_{n,i}(x) A_{m,j}(x, z)\, dT_x \qquad [A_{m,0} = K_m^0] \tag{7.4}$$

for (z) in $\bar{D}_{n-1}$ (if $\rho = \rho(z)$, >0, is suitably chosen). The functions in the kernel of (7.4) have the general properties indicated subsequent to (5.3).

In consequence of (7.3) and (7.4) one has

$$\delta_n \leqq \sum_{\lambda_0 \leqq \lambda_{n,i} < \lambda_q} \left| \sum_{j_1=0}^{m+1} \lambda_{n,i}^{j_1} \int_{D_n}^{(m)} u_{n,i}(z_1) A_{m,j_1}(z_1, x)\, dT_{z_1} \right| \cdot \left| \sum_{j_2=0}^{m+1} \lambda_{n,i}^{j_2} \int_{D_n}^{(m)} u_{n,i}(z_2) A_{m,j_2}(z_2, y)\, dT_{z_2} \right| \tag{7.5}$$

for (x) and (y) in $\bar{D}_{n-1}$. Here (as well as in (7.4)) D_n may be replaced by D. On letting λ^* denote the greater of the numbers 1, $|\lambda_0|$, $|\lambda_q|$ from (7.5) it is inferred that

$$\delta_n \leqq (\lambda^*)^{2m+2} \sum_{j_1, j_2} \left\{ \sum_i \left| \int_D^{(m)} A_{m,j_1}(z_1, x) u_{n,i}(z_1)\, dT_{z_1} \right| \cdot \left| \int_D^{(m)} A_{m,j_2}(z_2, y) u_{n,i}(z_2)\, dT_{z_2} \right| \right. \quad ((x), (y) \text{ in } \bar{D}_{n-1}).$$

Thus

$$\delta_n \leqq (\lambda^*)^{2m+2} \sum_{j_1, j_2} \left\{ \left[\sum_i \left| \int_D^{(m)} A_{m,j_1}(z_1, x) u_{n,i}(z_1)\, dT_{z_1} \right|^2 \right]^{\frac{1}{2}} \cdot \left[\sum_i \left| \int_D^{(m)} A_{m,j_2}(z_2, y) u_{n,i}(z_2)\, dT_{z_2} \right|^2 \right]^{\frac{1}{2}} \right\}$$

for (x), (y) in $\bar{D}_{n-1}$. On making use of Bessel's inequality we obtain

$$\delta_n \leqq (\lambda^*)^{2m+2} \sum_{j_1, j_2} \left[\int_D^{(m)} A_{m,j_2}^2(z_1, x)\, dT_{z_1} \right]^{\frac{1}{2}} \cdot \left[\int_D^{(m)} A_{m,j_2}^2(z_2, y)\, dT_{z_2} \right]^{\frac{1}{2}} = (\lambda^*)^{2m+2} a(x, y) \tag{7.7}$$

$((x), (y)$ in $\bar{D}_{n-1})$, with

$$a(x, y) = a(x \mid D) a(y \mid D), \tag{7.7a}$$

where $a(x \mid D)$ is defined by (6.18).

It is essential to note that the last member in (7.7) is independent of the mode of subdivision of the interval (λ_0, λ_q) and is independent of n.

By (7.7) and in view of the definition of δ_n, given in (7.3), one has

$$V_\alpha^\beta \theta_n(x, y \mid \lambda) \leqq (\lambda^*)^{2m+2} a(x, y) \tag{7.8}$$

for (x), (y) in D_{n-1}. *Here V_α^β denotes variation*, with respect to λ, on the interval (α, β); λ^* is the greatest of the numbers $|\alpha|, |\beta|, 1$. Also, since $\theta_n(x, y \mid 0) = 0$ we have

$$|\theta_n(x, y \mid \lambda)| = |\Delta_0^\lambda \theta_n|$$

so that, in view of (7.7),

$$|\theta_n(x, y \mid \lambda)| \leqq (\lambda')^{2m+2} a(x, y) \tag{7.9}$$

$((x), (y)$ in $\bar{D}_{n-1})$ where λ' is the greater of the numbers $1, |\lambda|$.

For any integer $r \geqq n$, corresponding to (7.4), we have

$$u_{r,i}(z) = \sum_{j=0}^{m+1} \lambda_{n,i}^j \int_{D_n}^{(m)} u_{r,j}(x) A_{m,j}(x, z)\, dT_x \tag{7.10}$$

for (z) in $\bar{D}_{n-1}$. This is a consequence of the fact that D_r $(r \geqq n)$ contains D_n and that for (x) in $D - \bar{D}_n$

$$A_{m,j}(x, z) = 0 \qquad \text{(for } (z) \text{ in } \bar{D}_{n-1}). \tag{7.11}$$

By a reasoning of the type employed subsequent to (7.4) for the purpose of derivation of (7.8) and (7.9) we now proceed from (7.10), (7.11) obtaining the inequalities

$$V_\alpha^\beta \theta_r(x, y \mid \lambda) \leqq (\lambda^*)^{2m+2} a(x, y), \tag{7.12}$$

$$|\theta_r(x, y \mid \lambda)| \leqq (\lambda')^{2m+2} a(x, y) \tag{7.13}$$

for (x), (y) in $\bar{D}_{n-1}$ and for $r = n + 1, \cdots$.

If the θ_r $(r \geqq n)$ had a property in the nature of equicontinuity in (x), (y), for (x), (y) in $\bar{D}_{n-1}$, then the "Compactness" theorem of Carleman[2] would be applicable. Under the existing conditions, however, a modification of the developments in (C_1 ; pp. 21–24), together with (7.12), (7.13), enables assertion of the following.

Given $n(> 0)$, there exists a subsequence

$$\theta_{r(n,1)}, \theta_{r(n,2)}, \cdots \tag{7.14}$$

such that the limit

$$\lim_j \theta_{r(n,j)} = \theta(x, y \mid \lambda)$$

exists in $\bar{D}_{n-1}$, for real λ; moreover,

$$V_\alpha^\beta \theta(x, y \mid \lambda) \leqq (\rho^*)^{2m+2} a(x, y), \tag{7.15}$$

$$|\theta(x, y \mid \lambda)| \leqq (\rho')^{2m+2} a(x, y) \tag{7.15a}$$

for (x) *and* (y) *in* $\bar{D}_{n-1}$, for all real λ, except perhaps for a denumerable infinity of values λ. Offhand, θ may depend on n; *whenever* (x) *or* (y) *is in* $D - \bar{D}_{n-1}$, *we define this function* θ *as zero.*

We select a subsequence (7.14) for $n = 2$, obtaining a limiting function, ${}_1\theta(x, y \mid r)$ for (x), (y) in $\bar{D}_1$. From this subsequence we choose another subsequence

$$\theta_{r_{3,1}}, \theta_{r_{3,2}}, \theta_{r_{3,3}}, \cdots \tag{7.16}$$

converging, for (x), (y) in $\bar{D}_2$, to a limiting function ${}_2\theta$, which is necessarily identical with ${}_1\theta$ for (x), (y) in $\bar{D}_1$.

We replace the sequence (7.16) by the sequence

$$\theta_{r(2,1)}, \theta_{r_{3,2}}, \theta_{r_{3,3}}, \theta_{r_{3,4}}, \cdots \tag{7.17}$$

which has the same limit as the sequence (7.16). From (7.17) we select a sequence

$$\theta_{r(2,1)}, \theta_{r_{3,2}}, \theta_{r_{4,3}}, \theta_{r_{4,4}}, \theta_{r_{4,5}}, \cdots$$

converging to ${}_3\theta$ in $\bar{D}_3$; clearly

$${}_3\theta = \begin{cases} {}_2\theta & \text{(for } (x), (y) \text{ in } \bar{D}_2), \\ {}_1\theta & \text{(for } (x), (y) \text{ in } \bar{D}_1). \end{cases}$$

Continuing this process of consecutive selections *one obtains a sequence*

$$\theta_{r(2,1)}, \theta_{r_{3,2}}, \theta_{r_{4,3}}, \theta_{r_{5,4}}, \cdots$$

converging to a limiting function

$$\theta(x, y \mid \lambda)$$

for (x), (y) in D, for all real λ, except perhaps for a denumerable infinity of values λ; θ is independent of n and satisfies the conditions stated in connection with (7.15), (7.15a), this being so for all (x), (y) in D.

DEFINITION 7.1. *A function* $\theta(x, y \mid \lambda)$ *obtained as described above will be termed a spectrum of* F, *associated with the sequence of boundary value problems* (7.1).

THEOREM 7.1. *Associated with the sequence of boundary value problems* (7.1) *there exists at least one spectrum* θ, *satisfying the conditions stated in connection with* (7.15), (7.15a) (*cf.* (7.7a)); *we also note the statement preceding Definition* 7.1.

In the remainder of this section *we shall state a number of properties of* $\theta(x, y \mid \lambda)$. These properties are analogous to those found, for functions designated as θ, by Carleman in Chapter I of (C_1). These results may be proved with the aid of the theorems referred to in (C_1 ; 7–24), as well as of Theorem 7.1. We shall omit the details of proof.

According to the preceding, there exists a sequence

$$\theta_{m_1}, \theta_{m_2}, \cdots \qquad (m_1 < m_2 < \cdots)$$

such that

$$\lim \theta_{m_\nu} = \theta \qquad ((x), (y) \text{ in } D; \text{ as } m_\nu \to \infty). \tag{7.19}$$

With (7.19) in view *we shall write* $n = m_\nu$. The following is asserted.

For almost all (x) in D

$$\psi(x,\lambda) = \int_D^{(m)} \theta(x, y \mid \lambda) h(y)\, dT_y = \lim \psi_n(x,\lambda),$$

(7.20)

$$\psi_n(x,\lambda) = \int_D^{(m)} \theta_n(x, y \mid \lambda) h(y)\, dT_y,$$

whenever $h(y) \subset L_2$ in D. If, in addition, $g \subset L_2$ in D, then

$$\int_D^{(m)} g(x)\psi(x,\lambda)\, dT_x = \lim \int_D^{(m)} g(x)\psi_n(x,\lambda)\, dT_x$$

$$= \lim \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda) g(x) h(y)\, dT_x\, dT_y;$$

the order of integration in

$$\int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda) g(x) h(y)\, dT_x\, dT_y \qquad (g, h \subset L_2 \text{ in } D) \tag{7.21}$$

is immaterial.

On using (7.17) (for θ_n) it is inferred that

$$V_\alpha^\beta \psi_n(x,\lambda), \qquad V_\alpha^\beta \psi(x,\lambda) \leqq a_0 \left[\int_D^{(m)} |h(y)|^2\, dT_y\right]^{\frac{1}{2}} a(x \mid D),$$

where $a_0 = a_0(\alpha, \beta)$ is finite for α, β finite. Moreover,

$$|\psi_n(x,\lambda)|, \qquad |\psi(x,\lambda)| \leqq a_1 a^2 \left[\int_D^{(m)} |h(y)|^2\, dT_y\right]^{\frac{1}{2}}, \tag{7.22}$$

where

$$a_1 = (\lambda')^{2m+2} \qquad [\lambda' = \text{max. } (1, |\lambda|]. \tag{7.23}$$

We have, whenever g, $h \subset L_2$ in D,

$$V_\alpha^\beta \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda) g(x) h(y)\, dT_x\, dT_y \tag{7.24}$$

and

$$\left|\int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda) g(x) h(y)\, dT_x\, dT_y\right|$$

$$\leqq \left[\int_D^{(m)} |g(x)|^2\, dT_x\right]^{\frac{1}{2}} \left[\int_D^{(m)} |h(y)|^2\, dT_y\right]^{\frac{1}{2}};$$

these inequalities will hold for $\theta(x, y \mid \lambda)$ as well.

When $\omega(\lambda)$ is continuous for $\lambda_0 \leqq \lambda \leqq \lambda_1$ one has

$$\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \theta(x, y \mid \lambda) = \lim \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \theta_n(x, y \mid \lambda),$$

$$\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \int_D^{(m)} h(y)\theta(x, y \mid \lambda)\, dT_y = \lim \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \int_D^{(m)} h(y)\theta_n(x, y \mid \lambda)\, dT_y,$$

$$\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)h(x)g(y)\, dT_x\, dT_y$$

(7.25)

$$= \lim \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda)h(x)g(y)\, dT_x\, dT_y$$

for (x), (y) in D, whenever h, $g \subset L_2$ in D.

Also

$$\int_D^{(m)} (hy)\left[\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \theta(x, y \mid \lambda)\right] dT_y = \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \left[\int_D^{(m)} \theta(x, y \mid \lambda)h(y)\, dT_y\right],$$

$$\int_D^{(m)} \int_D^{(m)} g(x)h(y)\left[\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \theta(x, y \mid \lambda)\right] dT_x\, dT_y$$

$$= \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\left[d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)g(x)h(y)\, dT_x\, dT_y\right],$$

$$\int_D^{(m)} g(x)\left\{\int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \left[\int_D^{(m)} \theta(x, y \mid \lambda)h(y)\, dT_y\right]\right\} dT_x$$

$$= \int_{\lambda_0}^{\lambda_1} \omega(\lambda)\, d_\lambda \left[\int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)g(x)h(y)\, dT_x\, dT_y\right].$$

The integral

$$\int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \tag{7.26}$$

converges for all $h \subset L_2$ (in D). There exists a function $\psi(x)$, $\subset L_2$ (in D), such that

$$\psi_l(x) = \int_{-l}^{l} d_\lambda \left[\int_D^{(m)} \theta(x, y \mid \lambda)h(y)\, dT_y\right] \to \psi(x)$$

(as $l \to +\infty$), convergence being in the mean square for (x) in D.

8. Developments on the basis of spectra

For a fixed $n = m_\nu$ consider the set of functions $u_{n,i}$ $(i = 1, 2, \cdots)$. Recalling the character of this sequence it is observed that the Parceval's equality for the $u_{n,i}$, expressed with the aid of a Stieltjes integral, is of the form

$$B_n = \int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \leqq \int_D^{(m)} |h(y)|^2\, dT_y.$$

Here h is any function, $\subset L_2$ in D, and θ_n is the function introduced in (7.2). With $0 < l < \infty$, we have

$$B_{n,l} = \int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \leqq \int_D^{(m)} |h(y)|^2\, dT_y. \tag{8.1}$$

This may be asserted in view of the fact that

$$\int_\alpha^\beta d_\lambda \int_D^{(m)} \int_D^{(m)} \theta_n(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \geqq 0$$

for all real α, β. In consequence of (7.25)

$$\lim_n B_{n,l} = \int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y .$$

Hence by (8.1)

$$\int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \leqq \int_D^{(m)} |\, h(y)\,|^2\, dT_y .$$

Thus, on letting $l \to \infty$, we obtain (note convergence of the integral (7.26))

$$\int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y \leqq \int_D^{(m)} |\, h(y)\,|^2\, dT_y , \tag{8.2}$$

which is the *generalized Bessel's inequality* related to our problem; this inequality is associated with a certain sequence of homogeneous boundary value problems.

It will be convenient to introduce the following Definition.

DEFINITION 8.1. *Suppose F is self adjoint. It will be said that F is closed with respect to a spectrum θ if*

$$\int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\bar{h}(x)h(y)\, dT_x\, dT_y = \int_D^{(m)} |\, h(y)\,|^2\, dT_y \tag{8.3}$$

for every $h \subset L_2$ (in D).

It is observed that (8.3) is a particular instance of (8.2) and constitutes a *generalized Parceval's relation.*

Suppose now F is closed with respect to a spectrum θ. Let h be real valued. Replacing $h(x)$ in (8.3) by $a(x) + b(x)$, where $a(x)$, $b(x) \subset L_2$ (in D), in consequence of the closed character of F we obtain

$$\begin{aligned} &\int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)a(x)b(y)\, dT_x\, dT_y \\ &\quad + \int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)b(x)a(y)\, dT_x\, dT_y = 2\int_D^{(m)} a(y)b(y)\, dT_y . \end{aligned} \tag{8.4}$$

Interchanging (x) and (y) in the second term of the first member above, making use of the symmetry of θ and taking note of the permissibility of the interchange of order of integration in the integral (7.21), it is concluded that this term is equal to the first integral displayed in (8.4). Accordingly

$$\int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)a(x)b(y)\, dT_x\, dT_y = \int_D^{(m)} a(y)b(y)\, dT_y . \tag{8.5}$$

Let C be a "cube" in the m-space, defined by inequalities

$$c_j \leqq x_j \leqq c_j + L \qquad (j = 1, \cdots m;\ L > 0). \tag{8.6}$$

We choose the point $(c) = (c_1, \cdots c_m)$ and L sufficiently great so that

$$D \subset C, \tag{8.6a}$$

supposing that D is bounded. We define θ, for (x) and (y) in C, by the relations

$$\begin{aligned} &\theta(x, y \mid \lambda) = 0 \qquad ((x) \text{ in } C - D;\ (y) \text{ in } C), \\ &\theta(x, y \mid \lambda) = 0 \qquad ((y) \text{ in } C - D;\ (x) \text{ in } C). \end{aligned} \tag{8.7}$$

One then may express (8.5) in the form

$$\int_{-\infty}^{\infty} d_\lambda \int_C^{(m)} \int_C^{(m)} \theta(x, y \mid \lambda) a(x) b(y)\, dT_x\, dT_y = \int_C^{(m)} a(y) b(y)\, dT_y, \tag{8.8}$$

where $a(x)$, $b(y)$ are any functions $\subset L_2$ in C, while

$$a(x) = 0 \qquad (\text{in } C - D). \tag{8.8a}$$

Substitute, in particular,

$$b(y) = b_t(y) = b(t_1, \cdots t_m\, ; y_1, \cdots y_m),$$

where the function $b_t(y)$ is defined, for (t) and (y) in C, by the relations

$$\begin{aligned} &b_t(y) = 1 \qquad (c_j \leqq y_j \leqq t_j\, ; j = 1, \cdots m), \\ &b_t(y) = 0 \qquad (\text{elsewhere}).^{18} \end{aligned} \tag{8.9}$$

From (8.8) we then obtain

$$\int_{-\infty}^{\infty} d_\lambda\, \sigma(t; \lambda) = \int_{y_1=c_1}^{t_1} \cdots \int_{y_m=c_m}^{t_m} a(y)\, dT_y \tag{8.10}$$

where, in view of (8.8a),

$$\sigma(t; \lambda) = \int_D^{(m)} a(x) \left\{ \int_{y_1=c_1}^{t_1} \cdots \int_{y_m=c_m}^{t_m} \theta(x, y \mid \lambda)\, dT_y \right\} dT_x. \tag{8.10a}$$

With the aid of (8.10) the following result is established.

THEOREM 8.1. *Suppose θ is a spectrum, with respect to which F is closed, and that D is bounded. There is then on hand the following generalized Fourier expansion*

$$a(t) = \frac{\partial^m}{\partial t_1 \cdots \partial t_m} \int_{-\infty}^{\infty} d_\lambda(t; \lambda) \tag{8.11}$$

for almost all (t) in D; this is valid for all $a \subset L_2$ in D; here $\delta(t; \lambda)$ is defined by (8.10a), *with θ subject to* (8.7) (*defined in* (8.6), (8.6a)).

An important application of spectra is, as will be now shown, in establishing explicit formulas for solutions of the non homogeneous problem (5.1).

We have

$$u_n(x) = \sum_i u^{n,i} u_{n,i}(x). \tag{8.12}$$

---

[18] That is, for points (y) such that at least some y_j exceeds t_j ((y), (t) in C).

Ordinary convergence, here, instead of convergence in the mean square takes place by reasoning of the same type as is used to prove the analogous fact in certain known simpler classical cases. As has been shown previously $u_n(x) \to u(x)$ (as $n = k_j \to \infty$) in the weak sense as well as in the ordinary sense. Certainly

$$\int_{(C)}^{(x)} u_{k_j}(x)\, dT_x \to \int_{(C)}^{(x)} u(x)\, dT_x \tag{8.13}$$

as $k_j \to \infty$, for (x) in C.[19] Here and in the sequel

$$\int_{(C)}^{(x)} a(x)\, dT_x = \int_{t_1=C_1}^{x_1} \cdots \int_{t_m=C_m}^{x_m} a(t_1, \cdots t_m)\, dt_1 \cdots dt_m \tag{8.14}$$

((x) in C; $a(x)$ integrable over C). Moreover, the functions $u_n(x)$ are taken equal to zero in $C - D$.

In view of (6.10) and of (6.11), from (8.12) it is deduced that

$$\begin{aligned} u_n(x) &= \sum_i \frac{1}{\lambda - \lambda_{n,i}} f^{n,i} u_{n,i}(x) = \sum_i \frac{1}{\lambda - \lambda_{n,i}} u_{n,i}(x) \int_D^{(m)} f(y) u_{n,i}(y)\, dT_y \\ &= \int_{-\infty}^{\infty} \frac{1}{\lambda - \rho}\, d_\rho \int_D^{(m)} f(y)\theta_n(x, y \mid \rho)\, dT_y . \end{aligned} \tag{8.15}$$

We write

$$u_n(x) = \tau_{n,l}(x) + r_{n,l}(x), \tag{8.16}$$

$$\tau_{n,l}(x) = \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho \int_D^{(m)} f(y)\theta_n(x, y \mid \rho)\, dT_y, \tag{8.16a}$$

$$r_{n,l}(x) = \left(\int_l^{\infty} + \int_{-\infty}^{-l}\right) \frac{1}{\lambda - \rho}\, d_\rho \int_D^{(m)} f(y)\theta_n(x, y \mid \rho)\, dT_y . \tag{8.16b}$$

On making use of the symbol (8.14) and employing the extended definition (8.7) of θ_n one may express (8.16)–(8.16b) in the form

$$\int_{(c)}^{(x)} u_n(x)\, dT_x = \int_{(c)}^{(x)} \tau_{n,l}(x)\, dT_x + \int_{(c)}^{(x)} r_{n,l}(x)\, dT_x, \tag{8.17}$$

$$\int_{(c)}^{(x)} \tau_{n,l}(x)\, dT_x = \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho \int_{(c)}^{(x)} \int_C^{(m)} f(y)\theta_n(x, y \mid \rho)\, dT_x\, dT_y, \tag{8.17a}$$

$$\int_{(c)}^{(x)} r_{n,l}(x)\, dT_x = \left(\int_l^{\infty} + \int_{-\infty}^{-l}\right) \frac{1}{\lambda - \rho}\, d_\rho \int_{(c)}^{(x)} \int_C^{(m)} f(y)\theta_n(x, y \mid \rho)\, dT_x dT_y . \tag{8.17b}$$

From (8.17a), (8.17b) we finally obtain

$$\int_{(c)}^{(x)} \tau_{n,l}(x)\, dT_x = \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y)\theta_n(t, y \mid \rho)\, dT_t\, dT_y, \tag{8.18}$$

$$\int_{(c)}^{(x)} r_{n,l}(x)\,dT_x$$

$$= \left(\int_l^{\infty} + \int_{-\infty}^{-l}\right) \frac{1}{\lambda - \rho}\, d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_n(t, y \mid \rho)\, dT_t\, dT_y \tag{8.19}$$

for (x) in C; here $b_x(t)$ is from (8.9).

Let T (Definition 6.1) *be of the form O* (cf. text subsequent to (6.14)). For λ in T one has

$$|\lambda - \lambda_{n,i}| \geqq \delta = \delta(\lambda) > 0 \qquad (n = k_1, k_2, \cdots; i = 1, 2, \cdots), \tag{8.20}$$

where $\delta(\lambda)$ is independent of n and i. Let $G(\delta)$ be the set of points whose distance from the interval $(-l, +l)$ is $<\delta$.

Case A. The point represented by λ is on the frontier of $G(\delta)$ or is exterior $G(\delta)$.

Case B. The point λ is in $G(\delta)$.

In Case A we have

$$|\lambda - \rho| \geqq \delta > 0 \qquad \text{(for all } \rho \text{ on } (-l, l)) \tag{8.21}$$

and, consequently,

$$\frac{1}{\lambda - \rho}, \tag{8.21a}$$

as a function of ρ, is continuous on the closed interval $(-l, l)$. In view of (7.25) from (8.18) it can be therefore deduced that

$$\lim_{k_j} \int_{(c)}^{(x)} \tau_{k_j,l}(x)\,dT_x = \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta(t, y \mid \rho)\, dT_t\, dT_y. \tag{8.22}$$

In Case B we take l sufficiently great so that the circle $S(\lambda, \delta)$, whose center is at λ and whose radius is δ, intersects the interval in two points, represented by numbers α, β, where

$$-l < \alpha < \beta < l.$$

The discussion, below, will still hold without any essential modifications when l is left unchanged. In view of (8.20) there are no points $\lambda_{n,i}$ $(n = k_1, k_2, \cdots; i = 1, 2, \cdots)$ in the open interval (α, β). Whence

$$\int_\alpha^\beta \frac{1}{\lambda - \rho}\, d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_n(t, y \mid \rho)\, dT_t\, dT_y = 0$$

$(n = k_1, k_2, \cdots)$ and, by (8.18),

$$\int_{(c)}^{(x)} \tau_{n,l}(x)\,dT_x = \left(\int_{-l}^{\alpha} + \int_\beta^l\right) \frac{1}{\lambda - \rho}$$

$$d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_n(t, y \mid \rho)\, dT_t\, dT_x. \tag{8.23}$$

In the field of integration indicated in (8.23) one has (8.21) and, accordingly, the function (8.21a) is continuous, in ρ, for ρ on the closed intervals

$$(-l, \alpha), \qquad (\beta, l).$$

We apply (7.25) to each of the integrals

$$\int_l^\alpha \cdots, \qquad \int_\beta^l \cdots,$$

letting $n = k_j \to \infty$. From (8.23) it is thus inferred that

$$\lim_{k_j} \int_{(c)}^{(x)} \tau_{k_j,l}(x)\, dT_x = \left(\int_l^\alpha + \int_\beta^l\right) \frac{1}{\lambda - \rho} \cdot \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta(t, y \mid \rho)\, dT_t\, dT_x. \tag{8.24}$$

On the other hand, inasmuch as

$$\int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_n(t, y \mid \rho)\, dT_t\, dT_x = \gamma_n(x) \qquad (n = k_1, k_2, \cdots),$$

where $\gamma_n(x)$ is independent of ρ for $\alpha < \rho < \beta$, and in so far as the limit

$$\lim_{k_j} \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_n(t, y \mid \rho)\, dT_t\, dT_y = \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta(t, y \mid \rho)\, dT_t\, dT_y = \gamma(x) \tag{8.25}$$

exists (in view of the developments of section 7), it is concluded that $\gamma(x)$ of (8.25) is independent of ρ for $\alpha < \rho < \beta$. The implication of the latter fact is the relation

$$\int_\alpha^\beta \frac{1}{\lambda - \rho}\, d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta(t, y \mid \rho)\, dT_t\, dT_y = 0. \tag{8.26}$$

By virtue of (8.26) one may put (8.24) in the form of (8.22); that is, *the equality* (8.22) *holds in any case* (for every λ in T).

In order to study the function (8.19) we take l so great that

$$-l < R\lambda < l \qquad (R\lambda = \text{real part of } \lambda).$$

It is then deduced that

$$|\lambda - \rho| \geqq l - R\lambda > 0, \tag{8.27}$$

for $\rho \geqq l$, and that

$$|\lambda - \rho| \geqq l + R\lambda > 0 \tag{8.27a}$$

for $\rho \leqq -l$. In view of (8.27) we find that

$$\left| \int_l^\infty \frac{1}{\lambda - \rho} d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_{k_j}(t, y \mid \rho)\, dT_t\, dT_y \right|$$

$$\leqq \frac{1}{l - R_\lambda} V_l^\infty \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_{k_j}(t, y \mid \rho)\, dT_t\, dT_y$$

$(j = 1, 2, \cdots)$. Now by virtue of (8.9) and (7.24)

$$V_l^\infty \cdots \leqq A = L^{\frac{1}{2}m} \left[\int_D^{(m)} |f(y)|^2\, dT_y \right]^{\frac{1}{2}},$$

where L is from (8.6); A is independent of k_j. Thus

$$\text{(8.28)} \quad \left| \int_l^\infty \frac{1}{\lambda - \rho} d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta_{k_j}(t, y \mid \rho)\, dT_t\, dT_y \right|$$

$$\leqq A \cdot (l - R\lambda)^{-1} \quad (j = 1, 2, \cdots; (x) \text{ in } C).$$

Similarly, on making use of (8.27a) it is concluded that

$$\text{(8.28a)} \quad \left| \int_{-\infty}^{-l} \text{integrand of the first member in (8.28)} \right| \leqq \frac{A}{l + R\lambda}$$

in C. Thus, by (8.28), (8.28a) and (8.19)

$$\text{(8.29)} \quad \left| \int_{(c)}^{(x)} r_{n,l}(x)\, dT_x \right| \leqq A \left[\frac{1}{l - R\lambda} + \frac{1}{l + R\lambda} \right] \qquad \text{(in } C)$$

for $n = k_1, k_2, \cdots$ and for λ in T. For any λ, fixed as specified, the second member in (8.29) can be made arbitrarily small by suitable choice of l; this can be done uniformly with respect to $n = k_j$ $(j = 1, 2, \cdots)$. Using this fact, as well as the relation (8.22) (valid in any case), in view of (8.17) it is deduced that

$$\text{(8.30)} \quad \lim_{k_j} \int_{(c)}^{(x)} u_{k_j}(x)\, dT_x = \int_{-\infty}^{\infty} \frac{1}{\lambda - \rho} d_\rho \int_C^{(m)} \int_C^{(m)} b_x(t) f(y) \theta(t, y \mid \rho)\, dT_t\, dT_y.$$

On taking account of (8.30), (8.9) and (8.13) it is inferred that

$$\text{(8.31)} \quad \int_{(c)}^{(x)} u(x)\, dT_x = \int_{-\infty}^{\infty} \frac{1}{\lambda - \rho} d_\rho \int_{(c)}^{(x)} \int_C^{(m)} f(y) \theta(x, y \mid \rho)\, dT_x\, dT_y.$$

Finally

$$\text{(8.32)} \quad u(x) = \frac{\partial^m}{\partial x_1 \cdots \partial x_m} \int_{-\infty}^{\infty} \frac{1}{\lambda - \rho} d_\rho \int_{(c)}^{(x)} \int_D^{(m)} f(y) \theta(x, y \mid \rho)\, dT_x\, dT_y$$

for (x) in D.

In view of the above it is possible to formulate the following Theorem.

THEOREM 8.2. *Suppose D is bounded. Every solution $u(x)$, for λ in O, referred to in Theorem* 6.2 *and satisfying the non homogeneous problem* (5.1), *is expressible in the form* (8.32) (*cf.* (8.14)).

The results of this section can be extended to hold for D unbounded if the integration $\int_{(c)}^{(x)}$ is replaced by integration $\int_e$, where e are Borel measurable subsets of D, with frontiers of zero measure, and if the derivation $\dfrac{\partial^m}{\partial x_1 \cdots \partial x_m}$ is replaced by set-function derivation D_x.

9. Some further developments involving spectra

We shall now essentially modify the procedures of section 8 and shall establish a number of spectral representations involving integrals convergent in the mean square. *Throughout, $\sim$ will denote convergence in the mean square.* We shall first establish the following result.

THEOREM 9.1. *Suppose F is closed with respect to a spectrum θ (Definition 8.1). Then every function $a(x)$, $\subset L_2$ in D, has a following representation*

$$a(x) \sim \int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) a(y)\, dT_y \qquad ((x) \text{ in } D). \tag{9.1}$$

To prove this we note that by virtue of the last statement of section 7 there exists a function $b(x)$, $\subset L_2$ in D, for which

$$b(x) \sim \int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) a(y)\, dT_y. \tag{9.2}$$

The function

$$\gamma(x) = b(x) - a(x)$$

belongs to L_2 ; in view of (8.3) and of the stated property of closure

$$\int_D^{(m)} \gamma^2(y)\, dT_y = \int_{-\infty}^{\infty} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\gamma(x)\gamma(y)\, dT_x dT_y = \lim_l w_l, \tag{9.3}$$

where

$$w_l = \int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\gamma(x)\gamma(y)\, dT_x dT_y. \tag{9.3a}$$

Replacing here $\gamma(y)$ by $b(y) - a(y)$ and making use of the formula preceding (7.26) one obtains

$$w_l = w_{l,1} - w_{l,2}, \tag{9.4}$$

where

$$w_{l,1} = \int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\gamma(x) b(y)\, dT_x dT_y$$

and

$$w_{l,2} = \int_{-l}^{l} d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda)\gamma(x) a(y)\, dT_x dT_y = \int_D^{(m)} \gamma(x) a_l(x)\, dT_x; \tag{9.5}$$

here

$$a_l(x) = \int_{-l}^{l} d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) a(y)\, dT_y.$$

In consequence of (8.5), which holds by virtue of the closure of F, one has

$$\lim_l w_{l,1} = \int_D^{(m)} \gamma(x) b(x)\, dT_x. \tag{9.6}$$

Now by (9.2)

$$\lim_l \int_D^{(m)} (b(x) - a_l(x))^2\, dT_x = 0.$$

Hence from (9.5) it is inferred that

$$\lim_l w_{l,2} = \int_D^{(m)} \gamma(x) b(x)\, dT_x. \tag{9.6a}$$

From (9.4), (9.6), (9.6a) we finally conclude that

$$\lim_l w_l = 0.$$

Whence, as may be seen from (9.3), $\gamma(y) = 0$; thus $b(x) = a(x)$ and the relation (9.2) becomes the representation (9.1), which was to be established.

THEOREM 9.2. *Let λ be in O. The problem*

$$F(u) + \lambda u = f \tag{9.7}$$

has a solution ($\subset L_2$ in D)

$$u(x) \sim \int_{-\infty}^{\infty} \frac{1}{\lambda - \rho}\, d_\rho \int_D^{(m)} f(y) \theta(x, y \mid \rho)\, dT_y. \tag{9.8}$$

It will be sufficient to give the proof for real λ in O and for f real valued. We shall write

$$s_n(x, \rho) = \int_D^{(m)} f(y) \theta_n(x, y \mid \rho)\, dT_y, \qquad s(x, \rho) = \int_D^{(m)} f(y) \theta(x, y \mid \rho)\, dT_y.$$

It will be first proved that, for $0 < l < +\infty$ and for λ in O,

$$\lim_{n_i} \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho s_{n_i}(x, \rho) = \int_{-l}^{l} \frac{1}{\lambda - \rho}\, d_\rho s(x \mid \rho). \tag{9.9}$$

With $\delta = \delta(\lambda)$ equal to the distance from λ to the closure of the set of points $\lambda_{n,i}$ $(n, i = 1, 2, \cdots)$ one has

$$d_\rho s_n(x \mid \rho) = 0$$

for

$$\lambda - \frac{\delta}{2} \leqq \rho \leqq \lambda + \frac{\delta}{2}$$

and

$$\int_l^l \frac{1}{\lambda-\rho}\,d_\rho s_n(x\,|\,\rho) = \left(\int_l^{\lambda'} + \int_{\lambda''}^l\right)\frac{1}{\lambda-\rho}\,d_\rho s_n(x\,|\,\rho) \tag{9.10}$$

$\left(\lambda' = \lambda - \frac{\delta}{2},\ \lambda'' = \lambda + \frac{\delta}{2};\ l \text{ suitably great}\right)$. By the formula preceding (7.25)

$$\lim_{n_i}\int_l^l \frac{1}{\lambda-\rho}\,d_\rho s_n(x\,|\,\rho) = \left(\int_l^{\lambda'} + \int_{\lambda''}^l\right)\frac{1}{\lambda-\rho}\,d_\rho s(x\,|\,\rho) \tag{9.10a}$$

$$= \int_l^l \frac{1}{\lambda-\rho}\,d_\rho s(x\,|\,\rho),$$

inasmuch as

$$d_\rho s(x\,|\,\rho) = 0 \qquad (\lambda' \leqq \rho \leqq \lambda'').$$

Thus (9.9) holds. We let

$$u_n(x, l) = \int_l^l \frac{1}{\lambda-\rho}\,d_\rho s_n(x\,|\,\rho),\quad r_n(x, l) = \left(\int_{-\infty}^{-l} + \int_l^{\infty}\right)\frac{1}{\lambda-\rho}\,d_\rho s_n(x\,|\,\rho). \tag{9.11}$$

Then, in view of (8.15),

$$u_n(x) = u_n(x, l) + r_n(x, l). \tag{9.11a}$$

It has been established previously that the limit

$$\lim u_{n_i}(x) = u(x)$$

exists in D; also,

$$\lim_{n_i} u_{n_i}(x, l) = u(x, l) = \int_l^l \frac{1}{\lambda-\rho}\,d_\rho s(x\,|\,\rho) \tag{9.11b}$$

in consequence of (9.9). Hence, by (9.11a), the limit

$$\lim_{n_i} r_{n_i}(x, l) = r(x, l) \tag{9.11c}$$

similarly exists. With the notation so introduced one has

$$u(x) = u(x, l) + r(x, l); \tag{9.11d}$$

also

$$\int_D^{(m)} |\,u(x) - u(x, l)\,|^2\,dT_x = \int_D^{(m)} |\,r(x, l)\,|^2\,dT_x. \tag{9.12}$$

We have (admitting complex values), by virtue of (9.11),

$$|\,r_n(x, l)\,|^2 = \sum_\alpha{}' \sum_\beta{}' \frac{1}{\lambda-\lambda_{n,\alpha}}\,\frac{1}{\bar\lambda-\lambda_{n,\beta}} \int_D^{(m)}\int_D^{(m)} f(y)\bar f(z)$$

$$\cdot u_{n,\alpha}(x)u_{n,\beta}(x)u_{n,\alpha}(y)u_{n,\beta}(z)\,dT_y\,dT_z,$$

where the primes over the summation symbols indicate that the sums are taken

corresponding to the intervals $(-\infty, -l)$, $(l, +\infty)$. Using the orthonormality relations of the $u_{n,i}$ $(i = 1, 2, \cdots)$ one accordingly obtains

$$\int_D |r_n(x, l)|^2 dT_x = \sum_\alpha{}' \frac{1}{|\lambda - \lambda_{n,\alpha}|^2} \int_D^{(m)} \int_D^{(m)} f(y)\bar{f}(z) u_{n,\alpha}(y) u_{n,\alpha}(z)\, dT_y\, dT_z.$$

The spectral form for this relation is

$$\int_D^{(m)} |r_n(x, l)|^2 dT_x = \left(\int_{-\infty}^{-l} + \int_l^{\infty}\right) \frac{1}{|\lambda - \rho|^2} d_\rho \psi_n(\rho),$$

$$\psi_n(\rho) = \int_D^{(m)} \int_D^{(m)} f(y)\bar{f}(z)\theta_n(y, z \mid \rho)\, dT_y\, dT_z.$$

In consequence of (7.24)

$$V_\alpha^\beta \psi_n(\rho) \leqq \int_D^{(m)} |f|^2 dT$$

for all finite α, β.

With $l > |R_\lambda|$ (λ real)

$$\left|\frac{1}{\lambda - \rho}\right| \leqq \frac{1}{l - R_\lambda} \quad (\rho \geqq l), \qquad \left|\frac{1}{\lambda - \rho}\right| \leqq \frac{1}{l + R_\lambda} \quad (\rho \leqq -l).$$

Then

$$\int_D^{(m)} |r_n(x, l)|^2 dT_x \leqq \left[\frac{1}{(l - R_\lambda)^2} + \frac{1}{(l + R_\lambda)^2}\right] \int_D^{(m)} |f|^2 dT = \sigma_\lambda(l).$$

The limiting function $r(x, l)$, of (9.11c), will satisfy the inequalities

$$\int_D^{(m)} |r(x, l)|^2 dT_x \leqq \varliminf_n \int_D^{(m)} |r_n(x, l)|^2 dT_x \leqq \sigma_\lambda(l).$$

Hence

$$\lim_{l \to \infty} \int_D^{(m)} |r(x, l)|^2 dT = 0.$$

Together with (9.12) and (9.11b), this implies that

$$\int_{-l}^{l} \frac{1}{\lambda - \rho} d_\rho s(x \mid \rho) \sim u(x) \qquad \text{(in } D\text{; as } l \to \infty\text{)},$$

which establishes (9.8) of the theorem.

We shall say that v is "*admissible*" if v is real, $\subset L_2$ (in D) and if

$$\frac{\partial v}{\partial \nu} + a_n(x) v = 0 \qquad (\text{on } S_n\,;\ n = n_1, n_2, \cdots) \tag{9.13}$$

for some functions $a_n(x)$, continuous on S_n $(n = 1, 2, \cdots)$.

It will be of interest to obtain *a spectral representation for the differential operator* F. We form the $u_{n,i}$, θ_n and θ, corresponding to the $a_n(x)$ $(n = n_1, n_2, \cdots)$ of (9.13). The following can be established (compare with (*C*)).

If v is "admissible" and $F(v) \subset L_2$ (in D), while F is closed with respect to θ, then

$$F(v(x)) \sim -\int_{-\infty}^{\infty} \lambda \, d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) v(y) \, dT_y \tag{9.14}$$

(when there is only one spectrum, it will be independent of v, of course); moreover, in order that the integral of the second member in (9.14) should converge in the mean square it is necessary and sufficient that the integral

$$\int_{-\infty}^{\infty} \lambda^2 \, d_\lambda \int_D^{(m)} \int_D^{(m)} \theta(x, y \mid \lambda) v(x) v(y) \, dT_x \, dT_y \tag{9.15}$$

should converge in the ordinary sense.

To prove (9.14) we apply the Fundamental Formula to v, $u_{n,i}$, obtaining the relation

$$\int_D^{(m)} F(v) u_{n,i} \, dT_x = -\lambda_{n,i} \int_D^{(m)} v u_{n,i} \, dT_x. \tag{9.16}$$

Now by Theorem 9.1

$$F(v(x)) \sim q_l(x) \qquad (\text{as } l \to +\infty), \tag{9.17}$$

where

$$q_l(x) = \int_{-l}^{l} d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) F(v(y)) \, dT_y.$$

In view of the formula preceding (7.25)

$$q_l(x) = \lim_n q_{l,n}(x) \qquad (\text{as } n = n_j \to \infty),$$

with

$$q_{l,n}(x) = \int_{-l}^{l} d_\lambda \int_D^{(m)} \theta_n(x, y \mid \lambda) F(v(y)) \, dT_y = \sum_i{}' u_{n,i}(x) \int_D^{(m)} F(v(y)) u_{n,i}(y) \, dT_y,$$

where the summation corresponds to the interval $(-l, l)$. By (9.16)

$$q_{l,n}(x) = -\sum_i{}' \lambda_{n,i} u_{n,i}(x) \int_D^{(m)} v(y) u_{n,i}(y) \, dT_y$$

$$= -\int_{-l}^{l} \lambda \, d_\lambda \int_D^{(m)} \theta_n(x, y \mid \lambda) v(y) \, dT_y.$$

In view of the relation preceding (7.25)

$$q_l(x) = -\int_{-l}^{l} \lambda \, d_\lambda \int_D^{(m)} \theta(x, y \mid \lambda) v(y) \, dT_y.$$

This, together with (9.17), implies (9.14). The statement with respect to (9.15) can be proved with the aid of section 7.

10. Operators A and their application

Let λ be in O and let u_n be solution of the problem

$$F(u_n) + \lambda u_n = f \qquad (\text{in } D_n),$$

$$\frac{\partial u_n}{\partial \nu} + a_n(x)u_n = 0 \qquad (\text{on } S_n),$$

while w_n is solution of the problem

$$F(w_n) + \lambda w_n = q \qquad (\text{in } D_n),$$

$$\frac{\partial w_n}{\partial \nu} + a_n(x)w_n = 0 \qquad (\text{on } S_n);$$

here f, q are continuous and

$$f, q \subset L_2 \qquad (\text{in } D). \tag{10.1}$$

It is to be noted that the set O is independent of the choice of f and q. *We define u_n, w_n as zero in $D - (D_n + S_n)$.*

For $u = u_n$ and $v = w_n$ application of the Fundamental Formula (5.11) will yield

$$\int_D^{(m)} w_n f \, dT_x = \int_D^{(m)} u_n q \, dT_x \qquad (u = n_1, n_2, \cdots). \tag{10.2}$$

There exists a subsequence $(k_1, k_2, \cdots)$ of $(n_1, n_2, \cdots)$ such that

$$w_{k_j}(x) \to w(x), \qquad u_{k_j}(x) \to u(x) \qquad (\text{as } k_j \to \infty; \text{ in } D), \tag{10.3}$$

where

$$u(x), w(x) \subset L_2 \qquad (\text{in } D),$$

while, for (x) in D, one has

$$F(u) + \lambda u = f, \tag{10.4}$$

$$F(w) + \lambda w = q; \tag{10.4a}$$

moreover,

$$\int_D^{(m)} |w_{k_j}(x)|^2 \, dT_x, \qquad \int_D^{(m)} |u_{k_j}(x)|^2 \, dT_x \leqq \sigma(\lambda) < +\infty \tag{10.5}$$

$(j = 1, 2, \cdots)$ with $\sigma(\lambda)$ independent of j. These facts are a consequence of the Existence Theorem 6.2.

In view of (10.5), (10.3) and (10.1) from (10.2) we derive

$$\int_D^{(m)} wf \, dT = \int_D^{(m)} ug \, dT. \tag{10.6}$$

By a reasoning of the type employed for an analogous purpose in (C) or using spectral representations one is able to choose the subsequence $(k_1, k_2, \cdots)$ independent of f and q.

The limits $u(x)$, $w(x)$ (cf. (10.3)), so obtained, can be represented as

$$u(x) = A_x(\lambda \mid f), \qquad w(x) = A_x(\lambda \mid q). \tag{10.7}$$

THEOREM 10.1. *Let O be the set defined subsequent to* (6.14). *Suppose F is self adjoint. There exist an operator*

$$A_x(\lambda \mid \cdots)$$

so that, whenever continuous $f, q \subset L_2$ in D, formula (10.7) *will represent solutions of equations* (10.4), (10.4a), *respectively, for λ in O. This operator satisfies the identity*

$$\int_D^{(m)} A_x(\lambda \mid q) f(x)\, dT_x = \int_D^{(m)} A_x(\lambda \mid f) q(x)\, dT_x, \tag{10.8}$$

which is valid for every λ in O and for all f, q of the described type. $A_x(\lambda \mid \cdots)$ may depend on the choice of the $a_n(x)$.

Theorem 10.1 is analogous to a result in (C_1 ; p. 56).

Let $f, q \subset L_2$ (in D) and

$$F(f), F(q) \subset L_2 \qquad \text{(in } D\text{)}. \tag{10.9); (10.10}$$

We then have

$$F(f) + \lambda f \equiv f^* \subset L_2 \qquad \text{(in } D\text{)}, \tag{10.11}$$

$$F(q) + \lambda q \equiv q^* \subset L_2 \qquad \text{(in } D\text{)}. \tag{10.11a}$$

Suppose F is of class I, with respect to O.

Application of (10.8) will enable inversion of (10.11), (10.11a) so as to yield

$$\int_D^{(m)} q(x) f^*(x)\, dT_x = \int_D^{(m)} f(x) q^*(x)\, dT_x. \tag{10.12}$$

Replacing f^*, q^* in (10.12) by the expressions from (10.11), (10.11a) one obtains

$$\int_D^{(m)} q(x) F(f)\, dT_x = \int_D^{(m)} f(x) F(q)\, dT_x. \tag{10.13}$$

The condition (10.13) is accordingly necessary in order that F be of Class I in O.

Conversely, *suppose now that* (10.13) *holds for all functions $f, q \subset L_2$ (in D) for which one has* (10.10). If F were not of class I (for λ in O) there would be on hand a function $\varphi(x) \subset L_2$ and a value λ_1 in O so that

$$F(\varphi) + \lambda_1 \varphi = 0, \tag{10.14}$$

$$\int_D^{(m)} |\varphi(x)|^2\, dT \neq 0. \tag{10.14a}$$

Now, let us define $q(x)$ as a solution, $\subset L_2$ (in D), of the equation

$$F(q) + \lambda_1 q = \bar{\varphi}. \tag{10.15}$$

By Theorem 6.2 such a solution $q(x)$ exists. Having assumed (10.13), we shall have in particular

$$\int_D^{(m)} q(x)F(\varphi)\,dT_x = \int_D^{(m)} \varphi(x)F(q)\,dT_x. \tag{10.16}$$

Replacing $F(q)$ by the expression obtained from (10.15), in consequence of (10.16) one obtains

$$\int_D^{(m)} |\,\varphi(x)\,|^2\,dT_x = \lambda_1 \int_D^{(m)} \varphi(x)q(x)\,dT_x + \int_D^{(m)} q(x)F(\varphi)dT_x.$$

Thus, by (10.14)

$$\int_D^{(m)} |\,\varphi(x)\,|^2\,dT_x = \int_D^{(m)} q(x)[F(\varphi) + \lambda_1\varphi]\,dT_x = 0. \tag{10.17}$$

Now (10.17) is contrary to (10.14a). Hence $F \subset I$ (in O). Accordingly the following Theorem has been established.

THEOREM 10.2. *Consider self adjoint operators F. In order that F be of class I (for λ in O), in accordance with Definition* 6.2, *it is necessary and sufficient that*

$$\int_D^{(m)} q(x)F(f)\,dT_x = \int_D^{(m)} f(x)F(q)\,dT_x$$

for all functions f, q belonging to L_2 (in D), for which $F(f)$, $F(q)$ belong to L_2 (in D).

On repeatedly using (10.8), it can be shown that the following holds.

THEOREM 10.3. *If for a fixed non real λ_1 the equation*

$$F(u) + \lambda u = 0 \qquad (\text{in } D) \tag{10.18}$$

has no solutions,[20] $\subset L_2$, *the same will be true for all non real λ. For non real values of λ the number of distinct*[21] *solutions, $\subset L_2$, of* (10.18) *is the same.*

The proof of this theorem will be omitted, as it may be given following closely the lines of proof of analogous results in (C_1 ; pp. 55, 58); it would be necessary, however, to use some of the previous developments of this section.

We shall now consider questions of uniqueness of solutions for λ in O, that is, when λ may be real. Such considerations would correspond to certain developments given by *Trjitzinsky*[4] for singular integral equations. The following will be proved.

THEOREM 10.4. 1°. *If for a fixed λ_1, in O, the equation* (10.18) *has no solutions, $\subset L_2$, the same will be true for all non real λ.*

2°. *The number m of distinct solutions, $\subset L_2$, of* (10.18) *for any λ fixed in O is equal to or is greater than the number n of distinct solutions for non real values λ.*

Consider part 1°. If λ_1 is non real the conclusion 1° follows by virtue of

---

[19] Apply theorem (C_1; pp. 132, 133).

[20] Here and in the sequel trivial, i.e. null, solutions, of homogeneous problems are disregarded.

[21] I.e., linearly independent.

Theorem 10.3. Take λ_1 real in O.[22] If 1° fails, there exists a non real value λ and a corresponding function u, $\subset L_2$, so that (10.18) holds and

$$\int_D^{(m)} |u|^2 dT \neq 0.$$

One has

$$F(u) + \lambda_1 u = (\lambda_1 - \lambda)u. \tag{10.19}$$

Now the equation

$$F(\varphi) + \lambda_1 \varphi = f_1 \qquad (f_1 \subset L_2 \text{ in } D)$$

cannot have two distinct solutions, since otherwise their difference ω would satisfy (10.18) for λ_1, while

$$\int_D^{(m)} |\omega|^2 dT \neq 0,$$

contrary to hypothesis. Thus,one may express the relation (10.19) in the form

$$u(x) = A_x(\lambda_1 \,|\, (\lambda_1 - \lambda)u). \tag{10.20}$$

Now by (10.19)

$$F(\bar{u}) + \lambda_1 \bar{u} = (\lambda_1 - \bar{\lambda})\bar{u}$$

and, in place of (10.20), one obtains

$$\bar{u}(x) = A_x(\lambda_1 \,|\, (\lambda_1 - \bar{\lambda})\bar{u}). \tag{10.20a}$$

In consequence of (10.8)

$$\int_D^{(m)} A_x(\lambda_1 \,|\, (\lambda_1 - \bar{\lambda})\bar{u})(\lambda_1 - \lambda)u \, dT_x = \int_D^{(m)} A_x(\lambda_1 \,|\, (\lambda_1 - \lambda)u)(\lambda_1 - \bar{\lambda})\bar{u} \, dT_x$$

and, substituting from (10.20), (10.20a), we deduce

$$(\lambda_1 - \lambda) \int_D^{(m)} |u|^2 dT_x = (\lambda_1 - \bar{\lambda}) \int_D^{(m)} |u|^2 dT_x$$

which implies that λ is real, contrary to our supposition.

We now proceed to part 2°. By Theorem 10.3 n is independent of λ. If 2° does not hold, there is a value λ, in O, for which the number m of distinct solutions,

$$u_1(x), \cdots u_m(x),$$

is less than n. By Theorem 10.3λ must be real. Let λ^* be a non real value; with λ^* there will be associated n distinct solutions,

$$u_1^*(x), \cdots u_n^*(x).$$

We have

$$F(u_j) + \lambda u_j = 0, \qquad F(u_\nu^*) + \lambda^* u_\nu^* = 0 \qquad \text{(in } D\text{)}.$$

[22] Supposing that O has real points; cf. text subsequent to (6.14).

Since λ is real the u_i may be taken real. Any solution u^* for λ^* is expressible in the form

$$u^*(x) = \sum_1^n c_\nu u_\nu^*(x) \qquad \text{(constants } c_\nu\text{)}.$$

With $m < n$, the c_ν may be chosen so that $u^*(x)$ will satisfy

$$\int_D^{(m)} |u^*(x)|^2\, dT \neq 0, \qquad \int_D^{(m)} u^*(x)\bar{u}(x)\, dT = 0 \tag{10.21}$$

for all solutions $u(x)$ for the value λ.

We have

$$F(u^*) + \lambda^* u^* = 0, \qquad F(\bar{u}^*) + \bar{\lambda}^* \bar{u}^* = 0;$$

whence

$$\begin{aligned} F(u^*) + \lambda u^* &= (\lambda - \lambda^*)u^* \equiv f, \\ F(\bar{u}^*) + \lambda \bar{u}^* &= (\lambda - \bar{\lambda}^*)\bar{u}^* \equiv \bar{f}. \end{aligned} \tag{10.22}$$

Accordingly

$$\begin{aligned} u^*(x) &= A_x(\lambda \mid f) + w_1(x), \\ \bar{u}^*(x) &= A_x(\lambda \mid \bar{f}) + w_2(x); \end{aligned} \tag{10.23}$$

here $w_1(x)$, $w_2(x)$ are some solutions of (10.18) for λ. Since the operator A can be so defined that

$$\bar{A}_x(\lambda \mid w) = A_x(\lambda \mid \overline{w}) \qquad \text{(all } w \subset L_2\text{)},$$

we have

$$w_2(x) = \overline{w}_1(x).$$

By (10.8)

$$\int_D^{(m)} A_x(\lambda \mid \bar{f}) f\, dT_x = \int_D^{(m)} A_x(\lambda \mid f)\bar{f}\, dT_x.$$

Substitution of (10.23) and of the expression for f, $\bar{f}$ from (10.22) will yield

$$(\lambda - \lambda^*) \int_D^{(m)} (\bar{u}^* - \overline{w}_1)u^*\, dT = (\lambda - \bar{\lambda}^*) \int_D^{(m)} (u^* - w_1)\bar{u}^*\, dT.$$

Further, in view of the property stated in connection with (10.21)

$$(\lambda - \lambda^*) \int_D^{(m)} |u^*|^2\, dT_x = (\lambda - \bar{\lambda}^*) \int_D^{(m)} |u^*|^2\, dT_x.$$

This leads to the conclusion that λ must be real; thus there arises a contradiction, which completes the proof of the Theorem. This can be extended to the case when the number of solutions could possibly be infinite.

By 1° of Theorem 10.4, if for a fixed λ_1, in O, the equation (10.18) has no solutions, $\subset L_2$ and distinct from zero, F will be of class I for all non real λ;

by the necessary part of Theorem 10.2 this would imply that the identity of that Theorem holds (for all f, q of stated type); in consequence of the sufficient part one then infers that $F \subset I$ in every set O. Hence we have the following: *if the equation*

$$F(u) + \lambda u = 0$$

has no solutions, $\subset L_2$ *in* D *and distinct from zero, for a value* λ_1 *in some set* O, *then the same will be true for all* λ *in every conceivable set* O.

In another paper the present author intends to obtain direct conditions on the $A_{i,j}$, B_i, C under which F is of class I, *the latter property being important in connection with questions of uniqueness and the closure of* θ.

UNIVERSITY OF ILLINOIS
AND
INSTITUTE FOR ADVANCED STUDY

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ON A THEOREM OF TANNAKA AND KREIN

BY S. BOCHNER

(Received June 13, 1941)

Let G denote an arbitrary group, C the class of almost periodic functions on G and C_0 the sub-class of all *finite* linear combinations of (irreducible) representation coefficients. With the norm

$$\| f \| = \sup_{x \in G} | f(x) |,$$

C is a Banach space, and C_0 is a dense but not closed subspace. Therefore a functional $L(f)$ which is additive, that is

$$L(af + bg) = aL(f) + bL(g), \tag{1}$$

need not be bounded on C_0 and hence need not be continuable onto all of C. However if $L(f)$ is positive, that is

$$L(f) \geqq 0 \text{ for } f \text{ (real and)} \geqq 0, \qquad f \in C_0 \tag{2}$$

then $| L(g) | \leqq 2 \| g \| \cdot L(1)$ and thus $L(f)$ is bounded and has a (unique) extension onto all of C which is again positive. By a recent theorem of M. Krein,[1] assumption (2) can be replaced by the weaker assumption:

$$L(| g |^2) \geqq 0 \quad \textit{for} \quad g \in C_0 . \tag{3}$$

Thus, *if $L(f)$ on C_0 has the properties* (1) *and* (3) *it also has property* (2). This theorem of Krein includes an earlier theorem of T. Tannaka[2] that any functional on C_0 having property (1) and the additional property

$$L(f_1 \cdot f_2) = L(f_1) \cdot L(f_2)$$

has an extension onto C itself.

The proof of Krein is based on ideas of N. Wiener and I. Gelfand which are extraneous to the problem, and we are going to give a new proof which stays wholly within the technique of uniform approximation to elements of C by elements of C_0.

We consider a complete set of irreducible representations

$$\{\varphi^{\rho}_{pq}(x)\}, \qquad x \in G.$$

The letter ρ designates an element of an index set of suitable potency, and for

[1] *On positive functionals on almost periodic functions*, Doklady Moscou *30*, pp. 9-12, (1941).

[2] *Über den Dualitätssatz der nicht-kommutativen topologischen Gruppen*, Tohoku J. Math., 45, pp. 1–12, (1938).

each ρ, p, $q = 1, \cdots, h = h(\rho)$, where $h(\rho)$ is the (finite) dimension of the ρ^{th} representation. Any element of C_0 can be written in the form

$$f(x) = \sum_{\rho,p,q} a^{\rho}_{pq}\varphi^{\rho}_{pq}(x), \tag{4}$$

where only a finite number of the constants a^{ρ}_{pq} are $\neq 0$. Introducing

$$\gamma^{\rho}_{pq} = L(\varphi^{\rho}_{pq}) \tag{5}$$

we have

$$L(f) = \sum_{\rho,p,q} a^{\rho}_{pq}\gamma^{\rho}_{pq}.$$

We call $L(f)$ a *special* functional if only a finite number of the constants (5) are $\neq 0$. In this case, since

$$|a^{\rho}_{pq}| \leqq \|f\|,$$

we have

$$|L(f)| \leqq \|f\| \cdot \sum_{\rho,p,q} |\gamma^{\rho}_{pq}|.$$

Thus $L(f)$ is bounded and has a continuous extension onto C. Furthermore every non-negative element of C is a uniform limit of squares of elements of C_0. Hence we obtain

LEMMA. *For any special $L(f)$, property* (3) *implies property* (2).

We will next consider a *non-negative* almost periodic function $\Lambda(t)$ and its Fourier coefficients

$$\lambda^{\rho}_{pq} = M_t\{\Lambda(t)\overline{\varphi^{\rho}_{pq}(t)}\}.$$

If the function $\Lambda(t)$ is a class function then

$$\lambda^{\rho}_{pq} = \lambda_{\rho}\delta_{pq}.$$

In this case, $\Lambda(t)$ is called a *weight-function*, and, if only a finite number of the values λ^{ρ}_{pq} is $\neq 0$, a *special* weight function.

Take any element (4) of C_0 which can be put in the form $f = |g|^2$, $g \in C_0$, and introduce the family of elements

$$f_y = f(xy^{-1}).$$

Since $f_y = |g(xy^{-1})|^2$, and

$$\varphi^{\rho}_{pq}(xy^{-1}) = \sum_{r=1}^{h} \varphi^{\rho}_{pr}(x)\overline{\varphi^{\rho}_{qr}(y)},$$

we see that (3) implies $L(f_y) \geqq 0$ and therefore $M_t\{\Lambda(t)L(f_t)\} \geqq 0$, the latter inequality being explicitly

$$\sum_{\rho,p,q} a^{\rho}_{pq}\lambda_{\rho}\gamma^{\rho}_{pq} \geqq 0. \tag{6}$$

Thus given the functional $L(f)$ and the weight functional $\Lambda(t)$ there exists a functional ΛL such that

$$\Lambda L(\varphi^{\rho}_{pq}) = \lambda_{\rho} \cdot \gamma^{\rho}_{pq}$$

and, what is decisive, if L has property (3) then so does ΛL. Now if $\Lambda(t)$ is a special weight function then ΛL is a special functional, and by our lemma, (6) holds for any $f \geqq 0$ belonging to C_0.

Now let f be any *fixed* non-negative element from C_0, and consider a sequence of weight functions, then

$$\sum_{\rho,p,q} a^{\rho}_{pq} \lambda^{(n)}_{\rho} \gamma^{\rho}_{pq} \geqq 0.$$

where ρ varies over a fixed *finite* index set, say $\rho = 1, \cdots, m$, which is independent of n, and $n = 1, 2, \cdots$. However, given any *finite* index set $\{\rho\}$ there exists a sequence of special weight functions such that for each ρ from the set, $\lim_{n\to\infty} \lambda^{(n)}_{\rho} = 1$.[3] Hence we obtain

$$\sum_{\rho,p,q} a^{\rho}_{pq} \gamma^{\rho}_{pq} \geqq 0$$

for each non-negative element of C_0, and this completes the proof of (2).

PRINCETON UNIVERSITY

[3] S. Bochner and J. von Neumann, *Almost periodic functions in groups*, Trans. Amer. Math. Soc., 37, pp. 21-50 (1935), especially pp. 37–40.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ON THE UNIFORM DISTRIBUTION OF THE ROOTS OF CERTAIN POLYNOMIALS

BY P. ERDÖS

(Received July 21, 1941)

Let

$$\begin{array}{c} x_1^{(1)} \\ x_1^{(2)} \quad x_2^{(2)} \\ \cdots\cdots\cdots\cdots \\ x_1^{(n)} \quad x_2^{(n)} \cdots x_n^{(n)} \end{array}$$

be a triangular matrix, where, for each n,

$$1 \geqq x_1^{(n)} > x_2^{(n)} > \cdots > x_n^{(n)} \geqq -1.$$

Since $x_i^{(n)}$ may be written in the form $x_i^{(n)} = \cos(\vartheta_i^{(n)})$, where $0 \leqq \vartheta_i^{(n)} \leqq \pi$, we may define another triangular matrix

$$\begin{array}{c} \vartheta_1^{(1)} \\ \vartheta_1^{(2)} \quad \vartheta_2^{(2)} \\ \cdots\cdots\cdots\cdots \\ \vartheta_1^{(n)} \quad \vartheta_2^{(n)} \cdots \vartheta_n^{(n)} \end{array}$$

with

$$0 \leqq \vartheta_1^{(n)} < \vartheta_2^{(n)} < \cdots < \vartheta_n^{(n)} \leqq \pi.$$

Put $\omega_n(x) = \prod(x - x_i)$.[1] Suppose $0 \leqq A < B \leqq \pi$. We denote by $N_n(A, B)$ the number of the ϑ_i in (A, B). Let $-1 \leqq a < b \leqq 1$. Then we denote by $M_n(a, b)$ the number of the x_i in (a, b). It does not matter whether the intervals (A, B) and (a, b) are open or closed.

In a previous paper[2] Turán and the author proved that if

$$|\omega_n(x)| < \frac{f(n)}{2^n}$$

then

$$N_n(A, B) = \frac{B - A}{\pi} n + O(n^{\frac{1}{2}}(\log f(n))^{\frac{1}{2}}).$$

[1] We omit the upper index n where there is no danger of confusion.

[2] *On the uniformly dense distribution of certain sequences of points*, Annals of Math. Vol. 41 (1940), pp. 162–173.

In another paper[3] we proved that if $|l_k^{(n)}(x)| < c_1$ then

$$N_n(A, B) = \frac{B-A}{\pi} n + O[(B-A)n]^{\frac{1}{2}+\epsilon}.$$

($l_k^{(n)}(x)$ denotes the fundamental polynomials, i.e. $l_k^{(n)}(x) = \omega(x)/[\omega'(x_k)(x-x_k)]$ is of degree $n-1$, and $l_k(x_k) = 1$, $l_k(x_i) = 0$, $i \neq k$.)

In the present paper we are going to improve these results. First we prove

THEOREM 1. *Put* $x_0 = -1$, $x_{n+1} = 1$, *and let*

$$(1) \qquad \max_{-1 \leq x \leq 1} |\omega_n(x)| < \frac{c_2}{2^n} \quad \textit{and} \quad \max_{x_k \leq x \leq x_{k+1}} |\omega_n(x)| > \frac{c_3}{2^n}, \qquad k = 0, 1 \cdots n.$$

Then

$$N_n(A, B) = \frac{B-A}{\pi} n + O[\log n(B-A)].$$

This result is the best possible.

Next we prove

THEOREM 2. *Let* $|l_k(x)| < c_4$; *then*

$$N_n(A, B) = \frac{B-A}{\pi} n + O[(\log n)(\log n(B-A))]$$

if $|l_k(x)| < n^{c_5}$, *then*

$$N_n(A, B) = \frac{B-A}{\pi} n + O[(\log n)^2].$$

Theorem 2 is also the best possible. Theorems 1 and 2 can be generalized to

THEOREM 3. *Let* $\omega(x)$ *be such that*

$$\frac{c_6 f(n)}{2^n} < \max_{x_k < x \leq x_{k+1}} |\omega_n(x)| < \frac{c_7 f(n)}{2^n} \qquad k = 0, 1, 2 \cdots n;$$

then

$$N_n(A, B) = \frac{B-A}{\pi} n + O[(\log n)(\log f(n))].$$

Similarly, if $|l_k(x)| < c_8 f(n)$ *then*

$$N_n(A, B) = \frac{B-A}{\pi} n + O[(\log n)(\log nf(n))].$$

To prove Theorem 1 we first have to prove two lemmas.

LEMMA 1. *Suppose that* (1) *holds; then*

$$(2) \qquad \frac{c_9}{n} < \vartheta_{k+1} - \vartheta_k < \frac{c_{10}}{n}, \qquad k = 0, 1 \cdots n.$$

[3] *On interpolation iii*, ibid. pp. 510–553.

PROOF. A theorem of M. Riesz states that if $h(x)$ is a polynomial of degree n which assumes its absolute maximum in $(-1, 1)$ at the point x_0, and if $y_1, \cdots, y_r$ are the roots of $h(y) = 0$ in the interval $(-1, 1)$, then $|\theta_i - \theta_0| \geqq \pi/2n$, where $\cos\theta_1 = y_i$ and $\cos\theta_0 = x_0$. Thus if x_0 lies between the roots y_i and y_{i+1}, then $\theta_{i+1} - \theta_i \geqq \pi/n$. Also if $\max_{y_i \leqq x \leqq y_{i+1}} h(x)$ assumes its smallest value for $i = k$ then

$$\theta_{k+1} - \theta_k \leqq \pi/n.$$

Suppose that (2) does not hold, for example assume that

$$\vartheta_{k+1} - \vartheta_k > r(n)/n,$$

where $\overline{\lim}\, r(n) = \infty$. Take $\epsilon > 0$, and define u and v by the relations: u and v are symmetric with respect to $(x_k + x_{k+1})/2$, and $\arccos u - \arccos v = \pi/n + \epsilon$. Consider the polynomial $\phi(x) = \omega(x)\cdot(x - u)\cdot(x - v)/(x - x_k)(x - x_{k+1})$. It can be seen that if $u \leqq x \leqq v$ then

$$\frac{(x-u)(x-v)}{(x-x_k)(x-x_{k+1})} < c_{11}/r(n);$$

hence

$$\max_{u \leqq x \leqq v} |\phi(x)| < (c_{11}/r(n)) \max_{x_k \leqq x \leqq x_{k+1}} \omega_n(x). \tag{3}$$

Also, since the sum of two quantities whose sum is fixed increases as they tend to equality, we have, in the intervals $(-1, x_k)$ and $(x_{k+1}, 1)$,

$$|\phi(x)| > |\omega(x)|. \tag{4}$$

We have $\arccos v - \arccos u > \pi/n$; and a simple calculation shows that, if $r(n)$ is large enough, $\vartheta_{k+1} - \arccos v > \pi/n$ and $\arccos u - \vartheta_k > \pi/n$; thus it follows from the lemma of M. Riesz (applied to $\phi(x)$) that $\max |\phi(x)|$ between two consecutive roots of $\phi(x)$, assumes its smallest value between the roots x_i and x_{i+1}, where either $i \leqq k - 2$ or $i \geqq k + 2$. Thus, from (3) and (4),

$$\max_{x_k \leqq x \leqq x_{k+1}} |\omega(x)| > \frac{r(n)}{c_{11}} \cdot \min_{j=0,1,\cdots,n} \max_{x_j \leqq x \leqq x_{j+1}} |\omega(x)|.$$

This contradicts (1), which completes the proof. By the same argument we could prove the other inequality in (2).

COROLLARY. *We obtain from Lemma* 1, *by a simple computation, that*

$$\frac{c_{12}}{n}\cdot(1 - x_k^2)^{\frac{1}{2}} < x_{k+1} - x_k < \frac{c_{13}}{n}(1 - x_k^2)^{\frac{1}{2}}, \qquad (k = 1, 2, \cdots, n-1).$$

LEMMA 2. *Suppose that* (1) *holds; then for* $-1 \leqq x \leqq 1$,

$$|l_k(x)| < c_{14}\frac{(1 - x_k^2)^{\frac{1}{2}}}{(x - x_k)n}.$$

PROOF. We have $l_k(x) = \omega(x)/[\omega'(x_k)(x - x_k)]$; thus by (1) it suffices to show that

$$\omega'(x_k) > c_{15}\frac{n}{2^n(1 - x_k^2)^{\frac{1}{2}}}.$$

Consider the polynomial $\psi(x) = \omega(x)/(x - x_k)$. It is clear that either $|\omega'(x_k)| = \psi(x_k) \geqq |\psi(y)|$ if $x_{k-1} \leqq y \leqq x_k$, or $|\omega'(x_k)| = \psi(x_k) \geqq \psi(y)$ if $x_k \leqq y \leqq x_{k+1}$. Without loss of generality we can assume that the first inequality holds. Then by (1) and the corollary to lemma 1 we have

$$\omega'(x_k) = \frac{\max_{x_{k-1} \leqq x \leqq x_k} |\omega(x)|}{x_k - x_{k-1}} > c_{15}\frac{n}{2^n(1 - x_k^2)^{\frac{1}{2}}},$$

which completes the proof.

Now we can prove Theorem 1. To simplify the calculations we assume that $a = 0$, $b = 1$. Then we have to show that, assuming (1)

$$\frac{n}{2} - c_{16} \log n < M_n(0, 1) < \frac{n}{2} + c_{17} \log n.$$

It will be sufficient to prove the first inequality. Suppose that it does not hold; then

$$M_n(0, 1) < \frac{n}{2} - r(n) \log n, \qquad \overline{\lim}\, r(n) = \infty.$$

Consider the polynomial $g(x)$ whose roots are defined as follows: In the interval $(-1, \log n/n)$, $g(x)$ has the same roots as $T_{n-1}(x)$ ($T_n(x)$ denotes the n^{th} Tchebicheff polynomial); at the points $(\frac{3}{2})^r \log n/n$, $r = 1, 2, \cdots s$ where s is such that

$$\left(\frac{3}{2}\right)^s \frac{\log n}{n} \leqq 1 < \left(\frac{3}{2}\right)^{s+1} \frac{\log n}{n},$$

$g(x)$ has a root of multiplicity $\left[\frac{r(n)}{10}\right]$; and finally $g(x)$ vanishes at the roots of $\omega(x)$ in the interval $(0, 1)$. Clearly the degree of $g(x)$ does not exceed

$$\frac{n}{2} + \log n + \frac{3 \log n}{10} r(n) + \frac{n}{2} - r(n) \log n < n - 1$$

if $r(n) > 10$. Thus, by the lemma of M. Riesz, $g(x)$ assumes its absolute maximum in the interval $(\log n/n, 1)$. Suppose that it assumes its absolute maximum at x_0, $\log n/n \leqq x_0 \leqq 1$. We have for some r

$$\left(\frac{3}{2}\right)^r \frac{\log n}{n} < x_0 < \left(\frac{3}{2}\right)^{r+1} \frac{\log n}{n}.$$

(If $(\frac{3}{2})^{r+1} \log n/n > 1$, we replace it by 1.) Put $(\frac{3}{2})^r \log n/n = q$; we consider the polynomial

$$g_1(x) = \frac{g(x)}{(x-q)^p}, \qquad p = \left[\frac{r(n)}{10}\right].$$

By the Lagrange interpolation formula we evidently have

$$g_1(x) = \sum_{k=1}^{n} g_1(x_k) l_k(x)$$

where the x_k are the roots of $\omega(x)$. Thus

$$g_1(x_0) = \sum_{k=1}^{n} g_1(x_k) l_k(x_0). \tag{5}$$

Now $g_1(x_k) = 0$ for $0 \leqq x \leqq 1$; and since x_0 was the place where $g(x)$ takes its absolute maximum, we have

$$g_1(x_0) \geqq [2(t+1)]^p g_1(x)$$

if x satisfies

$$-\frac{\log n}{n} t\left(\frac{3}{2}\right)^r \geqq x \geqq -\frac{\log n}{n}(t+1)\left(\frac{3}{2}\right)^r; \qquad t = 0, 1, 2 \cdots \tag{6}$$

(6) may be verified by noting that

$$g_1(x_0) = \frac{g(x_0)}{(x_0-q)^p} \geqq \frac{g(x)}{(x_0-q)^p} = g_1(x)\left(\frac{x-q}{x_0-q}\right)^p.$$

Hence from (5) and (6), by putting

$$-\frac{\log n}{n} t\left(\frac{3}{2}\right)^r = u_t,$$

we obtain

$$1 \leqq \sum_{t \geqq 0} \frac{M_n(u_t, u_{t+1}) \max_{u_t \geqq x_k \geqq x_{k+1}} |l_k(x_0)|}{2(t+1)^p} = \sum_1 + \sum_2$$

where in $\sum_1 t$ is restricted by $u_t \geqq -\frac{1}{2}$. Now by the corollary to Lemma 1, and Lemma 2.

$$\sum_1 < \sum_{t \geqq 0} c_{18} n(u_{t+1} - u_t) c_{19} \frac{1}{n_{t+1}} \frac{1}{[2(t+1)]^p} < c_{20} \sum_{t \geqq 0} \frac{1}{[2(t+1)]^p} < \frac{1}{2}$$

for sufficiently large p.

For the x_k in $\sum_2$ we clearly have $x_k < -\frac{1}{3}$. Thus by lemma 2.

$$\sum_2 < c_{21} \frac{n}{2^p} \max_{x_k < -\frac{1}{3}} |l_k(x_0)| < \frac{1}{2}$$

for sufficiently large p. Thus $\sum_1 + \sum_2 < 1$, and this contradiction establishes the proof.

In the proof we did not use the full strength of Lemma 2; in fact we only used

$|l_k(x)| < c_{22}\dfrac{1}{n|x - x_k|}$. We would have had to use the sharper estimate if we had not restricted ourselves to the interval (0, 1) but had considered a "small" interval near -1 or $+1$.

Now we have to prove that the error term in Theorem 1 is the best possible. Put

$$\vartheta_0 = \frac{\pi}{2}, \quad \vartheta_k = \frac{\pi}{2} + \frac{k\pi}{n} + \sum_{i=1}^{k}\frac{1}{i}, \qquad \vartheta_l = \frac{\pi}{2} - \frac{l\pi}{n} - \sum_{i=1}^{l}\frac{1}{i}$$

where k and l take all positive integral values such that $\vartheta_k < \pi - n^{-2}$, and $\vartheta_l > n^{-2}$ it is easy to see that the number of the ϑ's is $n + O(1)$. Consider the polynomial $\omega(x)$ whose roots are the $\cos \vartheta$'s. It can be shown by elementary computations that $\omega(x)$ satisfies (1). We do not give the details. On the other hand it is easy to see that

$$M_n(0, 1) < \frac{n}{2} - c_{23}\log n$$

which shows that the error term in Theorem 1 is the best possible.

The proof of Theorem 2 is very similar to that of Theorem 1. The difference is that, in defining $g(x)$, $g(x)$ now has roots of order $\left[\dfrac{r(n)\log n}{10}\right]$ at the points $(\frac{3}{2})^r \log n/n$. The proof of Theorem 3 also runs along the same lines.

UNIVERSITY OF PENNSYLVANIA

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ON THE ASYMPTOTIC DENSITY OF THE SUM OF TWO SEQUENCES

BY P. ERDÖS

(Received June 24, 1941)

Let $a_1 < a_2 < \cdots$ be an infinite sequence, A, of positive integers. Denote the number of a's not exceeding n by $f(n)$. Schnirelmann has defined the density of A as G.L.B. $f(n)/n$.[1] Now let $a_1 < a_2 < \cdots$; $b_1 < b_2 < \cdots$ be two sequences. We define the sum $A + B$ of these two sequences as the set of integers of the form a_i or b_j or $\{a_i + b_j\}$. Schnirelmann proved that if the density of A is α and that of B is β then the density of $A + B$ is $\geqq \alpha + \beta - \alpha\beta$.

Khintchine[2] proved that, provided that $\alpha = \beta \leqq \frac{1}{2}$, the density of $A + B$ is $\geqq 2\alpha$. He conjectured more generally that if $\alpha + \beta \leqq 1$ the density of $A + B$ is $\geqq \alpha + \beta$. It is easy to see that if $\alpha + \beta \geqq 1$ then every integer is in $A + B$, so the density of $A + B$ is 1. Khintchine's conjecture seems very deep.

Besicovitch[3] defined $\beta' =$ G.L.B. $\varphi(n)/(n + 1)$ where $\varphi(n)$ denotes the number of the b's not exceeding n, and proved that the Schnirelmann density of the sequence of numbers $\{a_i, a_i + b_j\}$ is $\geqq \alpha + \beta'$. An example of Rado showed that this result is the best possible.

Define the asymptotic density of A as $\underline{\lim}\, f(n)/n$. Then if $\alpha \leqq \frac{1}{2}$ and $a_1 = 1$ I have proved that the asymptotic density of $A + B$ is $\geqq \frac{3}{2}\alpha$.[4] The following simple example of Heilbronn shows that this result is the best possible: Let the a's be the integers $\equiv 0, 1 \pmod 4$. Then $A + A$ contains the integers $\equiv 0, 1, 2 \pmod 4$. In the present note we prove the following

THEOREM: *Let the asymptotic density of A be α and that of B be β, where $\alpha + \beta \leqq 1$, $\beta \leqq \alpha$, $b_1 = 1$. Then the asymptotic density of $A + B$ is not less than $\alpha + \frac{1}{2}\beta$, and, in fact, one of the sequences $\{a_i, a_i + 1\}$ or $\{a_i + b_j\}$ has asymptotic density $\geqq \alpha + \frac{1}{2}\beta$.*

It is easy to see that if $\alpha + \beta > 1$ then all large integers are in $A + B$. For if not then, none of the integers $n - a_i$ belong to B, and the asymptotic density of B would be not greater than $1 - \alpha < \beta$.

To prove our theorem we first need a slight sharpening of the theorem of Besicovitch; in fact, we prove the following

LEMMA: *Define the modified density of B as follows:*

---

[1] Schnirelmann, *Über additive Eigenschaften der Zahlen*, Math. Annalen 107 (1933), pp. 649–690.

[2] Khintchine, *Zur additiven Zahlentheorie*, Recueil math. de la soc. Moscow 39 (1932), pp. 27–34.

[3] Besicovitch, *On the density of the sum of two sequences of integers*, Journ. of the London math. soc. 10 (1935), pp. 246–248.

[4] Erdös, *On the asymptotic density of the sum of two sequences one of which forms a basis for the integers. ii.*, Travaux de l'institut math. de Tblissi 3 (1938), pp. 217–223.

1) $$\beta_1 = \underset{n>k}{\text{G.L.B.}} \frac{\varphi(n)}{n+1},$$

where the integers $1, 2, \cdots, k$ *belong to* B, *but* $k+1$ *does not belong to* B. *Clearly* $\beta_1 \geqq \beta'$. *Then the Schnirelmann density of the sequence* $\{a_i, a_i + b_j\}$ *is not less than* $\alpha + \beta_1$.

The proof of this lemma follows closely the proof of Besicovitch. Denote by $f(u, v)$, $\varphi(u, v)$, $\psi(u, v)$ respectively the number of a's, b's, and terms of the sequence $\{a_i, a_i + b_j\}$ in the interval (u, v)—that is, among the integers $u+1$, $u+2, \cdots, v$. We first observe that if $r+1$ is any integer which does not belong to the sequence $\{a_i, a_i + b_j\}$ then

2) $$f(u, v) + \varphi(r - v, r - u) \leqq v - u.$$

For as t runs through (u, v), $r + 1 - t$ runs through $(r - v, r - u)$, and if t belongs to A then $r + 1 - t$ does not belong to B.

We may assume that the Schnirelmann density of the sequence $\{a_i, a_i + b_j\}$ is less than 1, and that $\alpha > 0$, so that $a_1 = 1$. Define $m_0 = 0$, define $r_0 + 1$ as the least positive integer not belonging to $\{a_i, a_i + b_j\}$, define $m_1 + 1$ as the least integer greater than r_0 belonging to A, define $r_1 + 1$ as the least integer greater than m_1 not belonging to $\{a_i, a_i + b_j\}$, and so on.

It suffices to prove that for each x in (r_{i-1}, m_i) we have

3) $$\psi(0, x) \geqq (\alpha + \beta_1)x,$$

for if (3) holds, suppose that for some y in (m_j, r_j) we had

$$\psi(0, y) < (\alpha + \beta_1)y.$$

(We may suppose $j > 0$; else $y \leqq r_0$, so that $\psi(0, y) = y$). Then since all the integers $m_j + 1, \cdots, y$ belong to $\{a_i, a_i + b_j\}$ and $\alpha + \beta_1 \leqq 1$ we should have

$$\psi(m_j) < (\alpha + \beta_1)m_j,$$

which contradicts (3).

It follows from the definition of k and the definition of m_i and r_i that

4) $$r_i - m_i > k \qquad (i = 0, 1, 2 \cdots).$$

Let $r_{i-1} < x \leqq m_i$; we have

5) $$\psi(r_{i-1}, x) \geqq \varphi(r_{i-1} - m_{i-1} - 1, x - m_{i-1} - 1),$$

since any number $m_{i-1} + 1 + u$, where u belongs to B, is in $\{a_i, a_i + b_j\}$. Also

6) $$\psi(m_{i-1}, r_{i-1}) = r_{i-1} - m_{i-1} \geqq f(m_{i-1}, r_{i-1}) + \varphi(0, r_{i-1} - m_{i-1})$$

by (2). Clearly by the definition of the numbers r_j, m_j we have for $r_{i-1} < x \leqq m_i$, $f(m_{i-1}, x) = f(m_{i-1}, r_{i-1})$. Hence by adding (5) and (6)

7) $$\psi(m_{i-1}, x) \geqq f(m_{i-1}, x) + \varphi(0, x - m_{i-1} - 1) \geqq f(m_{i-1}, x) + \beta_1(x - m_{i-1}),$$

since by (4) $x - m_{i-1} - 1 \geqq r_{i-1} - m_{i-1} > k$. In particular

8) $\psi(m_j, m_{j+1}) \geqq f(m_j, m_{j+1}) + \beta_1(m_{j+1} - m_j) \qquad (j = 0, 1, \cdots).$

Summing (8) for $j = 0, 1, \cdots i - 1$ and adding (7) we have

$$\psi(0, x) \geqq f(0, x) + \beta_1 x \geqq (\alpha + \beta_1)x,$$

which completes the proof of the Lemma.

Now we can prove our theorem. We may assume $\beta > 0$. Suppose first that there exists an x belonging to A, such that the modified density of (the positive terms of) $a_i - x$ is $\geqq \alpha - \frac{1}{2}\beta$. Clearly $x + 1$ has to be in A since $\alpha - \frac{1}{2}\beta > 0$. It follows that there exists for every positive real ϵ a y such that the Schnirelmann density of the positive terms of the sequence $\{b_j - y\}$ is $\geqq \beta - \epsilon$. To see this choose y to be the greatest integer with

$$\frac{\varphi(y)}{y} \leqq \beta - \epsilon.$$

(Since $\underline{\lim}\, \varphi(y)/y = \beta$ such a y exists, unless $\varphi(y)/y > \beta - \epsilon$ for all positive y; in this case we have $y = 0$). Then by the definition of y it is clear that $\varphi(y, z)$ i.e. the number of $\{b_j - y\}$'s in $(0, z - y)$, is not less than $(\beta - \epsilon)(z - y)$, which proves our assertion.

Now consider the sequence $\{b_j - y, b_j - y + a_i - x\}$. By our lemma its Schnirelmann density is $\geqq \alpha + \frac{1}{2}\beta - \epsilon$; hence by adding $x + y$ to its members we obtain the sequence $\{b_j + x, a_i + b_j\}$ whose asymptotic density is clearly $\geqq \alpha + \frac{1}{2}\beta - \epsilon$ for every $\epsilon > 0$. But since x is in A, $b_j + x$ is in $\{a_i + b_j\}$. Hence the asymptotic density of the sequence $\{a_i + b_j\}$ is $\geqq \alpha + \frac{1}{2}\beta$, which proves our theorem in the first case.

Suppose next that Case 1 is not satisfied. We may suppose that there exist arbitrarily large values of i such that a_i and $a_i + 1$ are both in A; otherwise $\{a_i, a_i + 1\}$ has asymptotic density $2\alpha > \alpha + \frac{1}{2}\beta$. Let a_{k_1} be the first a_i such that $a_{k_1} + 1$ is also in A. Then since Case 1 is not satisfied and since $\alpha = \underline{\lim}\, f(n)/n$, there exists a largest integer m_1 such that $f(a_{k_1}, m_1) < (\alpha - \frac{1}{2}\beta)(m_1 - a_{k_1} + 1)$. Again let a_{k_2} be the least a_i greater than m_1 such that $a_{k_2} + 1$ is also in A; there exists as before a largest m_2 such that $f(a_{k_2}, m_2) < (\alpha - \frac{1}{2}\beta)(m_2 - a_{k_2} + 1)$ and so on. Take n large and let m_r be the least $m \geqq n$. It is clear that the intervals $(a_{k_i} - 1, m_i)$, $i = 1, 2 \cdots r$ do not overlap; thus

$$\sum_{i=1}^{r} f(a_{k_i}, m_i) \leqq m_r\left(\alpha - \frac{\beta}{2}\right).$$

Now since the asymptotic density of A is α, we have $f(0, m_r) > (\alpha - \epsilon)m_r$, if n is large enough, and therefore the number of a_i's in $(0, n)$ outside the intervals (a_{k_i}, m_i), $i = 1, 2 \cdots r$ is not less than

$$\left(\frac{\beta}{2} - \epsilon\right)m_r \geqq \left(\frac{\beta}{2} - \epsilon\right)n.$$

But for all these a_i's with the exception of $a_{k_1}, a_{k_2}, \cdots, a_{k_r}, a + 1$ is not in A.

Moreover, the intervals (a_{k_i}, m_i) do not contain only a's; else, whenever $p > a_{k_i}$ is such that (a_{k_i}, p) does contain integers not in A, we have $p > m_i$. Therefore $f(a_{k_i}, p) \geqq (\alpha - \frac{1}{2}\beta)(p - a_{k_i} + 1)$ (by definition of m_i); so that the modified density of the positive terms of $\{a_j - a_{k_i}\}$ $(j = 1, 2 \cdots)$ is $\geqq \alpha - \frac{1}{2}\beta$, and we are in Case 1. Thus each of the intervals (a_{k_i}, m_i) has to contain an x which is in A, such that $x + 1$ is not in A. Hence, finally, the number of integers $\leqq n$ of the form $a_i + 1$ which are not in A is $\geqq (\frac{1}{2}\beta - \epsilon)(n - 1)$. Hence the number of integers $\leqq n$ of the form $\{a_i, a_i + 1\}$ is not less than $(\alpha + \frac{1}{2}\beta - \epsilon)n - 1$, if n is large enough, which completes the proof of our theorem.

UNIVERSITY OF PENNSYLVANIA

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

SOME NEW SUMMABILITY METHODS WITH APPLICATIONS

BY OTTO SZÁSZ[1]

(Received November 4, 1941)

1. Given an infinite sequence of functions

$$\varphi_0(x), \varphi_1(x), \varphi_2(x), \cdots, \varphi_\nu(x), \cdots, \tag{1.1}$$

defined on a finite or infinite range R of the real or complex variable x with the limit point ξ; the range R may be discrete or continuous, but must contain infinitely many points. The *series-to-function transform*

$$\Phi(x) = \sum_{\nu=0}^{\infty} u_\nu \varphi_\nu(x) \tag{1.2}$$

defines, under certain assumptions for (1.1), a summability method, and $\lim_{x\to\xi} \Phi(x)$, if it exists, is called the generalized sum of the series $\sum_0^\infty u_\nu$. If it is true that for every convergent series $\sum u_\nu$, $\Phi(x)$ exists in R and $\lim \Phi(x) = \sum u_\nu$, then the method is called regular. The necessary and sufficient conditions for regularity are:

$$\lim_{x\to\xi} \varphi_\nu(x) = 1, \qquad \text{for } \nu = 0, 1, 2, \cdots \tag{1.3}$$

$$\sum_0^\infty | \varphi_\nu(x) - \varphi_{\nu+1}(x) | \text{ uniformly bounded on } R. \tag{1.4}$$

We then say that (1.2) is a regular transform; the summability method defined by

$$\lim_{x\to\xi} \Phi(x) = s \tag{1.5}$$

is called the method of convergence factors (according to C. N. Moore).[2] Corresponding to this method we obtain new methods in the following way:

We select a sequence of values $x = x_n \to \xi$, and associate with (1.2) a series-to-sequence transform

$$A_n(x_n) = A_n = \sum_{\nu=0}^{n} u_\nu \varphi_\nu(x_n), \qquad n = 0, 1, 2, \cdots; \tag{1.6}$$

its matrix is of triangular type. The summability method defined by

[1] Presented to the American Mathematical Society, September 2, 1941.

[2] For a comprehensive discussion of the general regular transforms cf. Szász [8], Agnew [1]; particular methods of the type (1.5) have been discussed by Perron [3]. Numbers in brackets refer to the literature listed at the end of this paper.

$$\lim_{n\to\infty} A_n = s \tag{1.7}$$

is regular (by the theorem of Silverman and Toeplitz) if and only if:

$$\lim_{n\to\infty} \varphi_\nu(x_n) = 1 \qquad \text{for } \nu = 0, 1, 2, \cdots, \tag{1.8}$$

$$\sum_{\nu=0}^{n-1} |\, \varphi_\nu(x_n) - \varphi_{\nu+1}(x_n) \,| + |\, \varphi_n(x_n) \,| \text{ is uniformly bounded in } n. \tag{1.9}$$

It is obvious from our regularity conditions that the method (1.7) is regular whenever (1.5) is, but the converse is not true in general.

We can apply the same generating process to a *sequence-to-function transform*:

$$\Psi(x) = \sum_{\nu=0}^{\infty} s_\nu \psi_\nu(x)\,, \tag{1.10}$$

where $\{\psi_\nu(x)\}$ is a suitably given sequence of functions. From (1.10) we get the summability method:

$$\lim_{x\to\xi} \Psi(x) = s = \text{gen. lim. } s_n. \tag{1.11}$$

This method is regular (i.e. $\lim s_n = s$ implies (1.11)) if and only if:

$$\lim_{x\to\xi} \psi_\nu(x) = 0, \qquad \text{for } \nu = 0, 1, 2, \cdots, \tag{1.12}$$

$$\sum_{\nu=0}^{\infty} |\, \psi_\nu(x) \,| \text{ uniformly bounded on } R, \tag{1.13}$$

$$\sum_{\nu=0}^{\infty} \psi_\nu(x) \to 1 \qquad \text{as } x \to \xi. \tag{1.14}$$

We now associate with (1.10) and with a sequence of values $x = x_n \to \xi$ a triangular-type sequence to sequence transform:

$$B_n(x_n) = B_n = \sum_{\nu=0}^{n} s_\nu \psi_\nu(x_n), \qquad n = 0, 1, 2, \cdots. \tag{1.15}$$

The summability method

$$\lim B_n = s = \text{gen. lim } s_n \tag{1.16}$$

is regular if and only if:

$$\lim_{n\to\infty} \psi_\nu(x_n) = 0, \qquad \text{for } \nu = 0, 1, 2, \cdots, \tag{1.17}$$

$$\sum_{\nu=0}^{n} |\, \psi_\nu(x_n) \,| \text{ uniformly bounded in } n, \tag{1.18}$$

$$\lim_{n\to\infty} \sum_{\nu=0}^{n} \psi_\nu(x_n) = 1. \tag{1.19}$$

It is seen easily that regularity of the method (1.11) does not imply regularity of the method (1.15), and conversely.

Note that A_n can also be written as a sequence to sequence transform:

$$A_n = \sum_{\nu=0}^{n-1} s_\nu\{\varphi_\nu(x_n) - \varphi_{\nu+1}(x_n)\} + s_n\varphi_n(x_n),$$

where

$$s_n = \sum_0^n u_\nu, \qquad n = 0, 1, 2, \cdots.$$

A generalization of (1.6) is

$$A_m(x_n) = \sum_{\nu=0}^{m} u_\nu\varphi_\nu(x_n), \qquad \text{where } m = m(n) \to \infty,$$

and the analogous sequence-to-function transform

$$F(x) = \sum_{\nu \leqq m(x)} u_\nu\varphi_\nu(x).$$

These generalizations will not be discussed in the present paper.

2. We consider first the important case of Abel-summability (generalized by Stolz to complex values of x); it can be written either as a convergence-factor method:

$$\lim_{x\to 1} \sum_0^\infty u_\nu x^\nu = s,$$

or as a sequence-to-function transform:

$$\lim_{x\to 1} \sum_0^\infty s_\nu x^\nu(1 - x) = s.$$

Accordingly:

(2.1) $$\varphi_n(x) = x^n, \qquad n = 0, 1, 2, \cdots,$$

and

(2.2) $$\psi_n(x) = (1 - x)x^n, \qquad n = 0, 1, 2, \cdots.$$

The method is known to be regular if and only if x approaches 1 along a path inside and non-tangential to the unit circle, i.e.

$$|x| < 1 \quad \text{and} \quad 1 - x = O(1 - |x|) \qquad \text{as } x \to 1.$$

Writing $x = \rho e^{i\theta}$ the latter condition becomes $\theta = O(1 - \rho)$.

Now the associated series to sequence transform is:

$$A_n = \sum_{\nu=0}^{n} u_\nu x_n^\nu = \sum_{\nu=0}^{n-1} s_\nu x_n^\nu (1 - x_n) + s_n x_n^n, \qquad n = 0, 1, 2, \cdots. \tag{2.3}$$

The regularity conditions are (from (1.8) and (1.9)):

$$x_n \to 1 \quad \text{as} \quad n \to \infty, \tag{2.4}$$

$$|x_n|^n + \frac{1 - |x_n|^n}{1 - |x_n|} |1 - x_n| = O(1) \qquad \text{as } n \to \infty.^3 \tag{2.5}$$

On the other hand, using (2.2) we get the sequence-to-sequence transform

$$B_n = \sum_{\nu=0}^{n} s_\nu x_n^\nu (1 - x_n), \tag{2.6}$$

and the summability method

$$\lim_{n\to\infty} B_n = s = \text{gen. lim } s_n,$$

which is regular if and only if:

$$x_n \to 1,\ x_n^n \to 0, \quad \text{and} \quad 1 - x_n = O(1 - |x_n|) \qquad \text{as } n \to \infty. \tag{2.7}$$

Although the methods $\lim A_n$ and $\lim B_n$ both originate from Abel's method in much the same way, they have quite different regularity conditions.

On writing

$$x_n = \rho_n e^{i\theta_n}, \qquad -\pi < \theta_n \leqq \pi,$$

the conditions (2.4) and (2.5) take the form

$$\rho_n \to 1, \qquad \theta_n \to 0, \qquad \rho_n^n = O(1), \tag{2.8}$$

and

$$1 - 2\rho_n \cos\theta_n + \rho_n^2 = O\left(\left(\frac{1 - \rho_n}{1 - \rho_n^n}\right)^2\right) \qquad \text{as } n \to \infty. \tag{2.9}$$

But

$$1 - 2\rho_n \cos\theta_n + \rho_n^2 = (1 - \rho_n)^2 + 4\rho_n \sin^2 \tfrac{1}{2}\theta_n,$$

and in view of (2.8)

$$1 - \rho_n = O\left(\frac{1 - \rho_n}{1 - \rho_n^n}\right) \qquad \text{as } n \to \infty,$$

and

$$\sin \tfrac{1}{2}\theta_n \sim \tfrac{1}{2}\theta_n,$$

hence (2.9) reduces to

$$\theta_n = O\left(\frac{1 - \rho_n}{1 - \rho_n^n}\right) \qquad \text{as } n \to \infty.$$

---

[3] If $|x_n| = 1$, then $\dfrac{1 - |x_n|^n}{1 - |x_n|}$ is to be replaced by n.

Note that

$$\frac{1-\rho_n}{1-\rho_n^n} = \frac{1}{1+\rho_n+\cdots+\rho_n^{n-1}} = o(1) \qquad \text{as } n \to \infty.$$

Finally, putting $\rho_n = 1 + \delta_n$, we have $\delta_n \to 0$, and $\rho_n^n = e^{n \log (1+\delta_n)}$; but

$$\log (1+\delta) = \delta(1 + o(1)) \qquad \text{as } \delta \to 0. \tag{2.10}$$

Hence $\rho_n^n = O(1)$ if and only if $n(\rho_n - 1) < k$.[4] Summarizing we have

THEOREM 1. *Necessary and sufficient conditions for the regularity of the transform* (2.3) (*where* $x_n = \rho_n e^{i\theta_n}$) *are*:

$$\rho_n \to 1, \quad \overline{\lim}\, n(\rho_n - 1) < +\infty, \quad \theta_n = O\left(\frac{1-\rho_n}{1-\rho_n^n}\right), \qquad \text{as } n \to \infty. \tag{2.11}$$

The last condition is certainly satisfied if $\theta_n = O(1/n)$; this condition is also a necessary one if $\rho_n \geqq 1$ for all large n.

For the regularity of the method (2.6) the necessary and sufficient conditions are (from (2.7))

$$\rho_n \to 1, \ \rho_n^n \to 0 \quad \text{and} \quad \theta_n = O(1-\rho_n) \qquad \text{as } n \to \infty;$$

here the second condition can be replaced by (using (2.10))

$$n(1-\rho_n) \to +\infty \qquad \text{as } n \to \infty.$$

It is easily seen that the last condition of (2.11) can be replaced by $\theta_n = O(n^{-1} + |1 - \rho_n|)$.

3. We get a summability method related to Abel's on choosing

$$\varphi_n(x) = \Re x^n = \rho^n \cos n\theta, \qquad n = 0, 1, 2, \cdots.$$

The regularity conditions (1.3) and (1.4) now become:

$$\rho \to 1, \qquad \theta \to 0 \quad \text{and}$$

$$\sum_{\nu=0}^{\infty} \rho^\nu \,|\, \cos \nu\theta - \rho \cos (\nu+1)\theta \,| \text{ uniformly bounded in } \rho \text{ and } \theta. \tag{3.1}$$

In particular we must have $\rho < 1$; furthermore, using the formulae

$$\cos \nu\theta - \rho \cos (\nu+1)\theta = (1-\rho)\cos \nu\theta + 2\rho \sin \tfrac{1}{2}\theta \sin (\nu + \tfrac{1}{2})\theta, \tag{3.2}$$

and

$$\sum_0^\infty \rho^\nu (1-\rho) \,|\, \cos \nu\theta \,| \leqq 1,$$

(3.1) is equivalent to:

$$|\, \theta \,| \sum_0^\infty \rho^\nu \,|\, \sin (\nu + \tfrac{1}{2})\,\theta \,| \text{ uniformly bounded in } \rho \text{ and } \theta.$$

[4] $k, k_1, k_2, \cdots$ denote absolute constants.

Thus a sufficient condition is:

$$\frac{\theta}{1-\rho} = O(1) \quad \text{for } \rho \to 1, \text{ or for } \theta \to 0.$$

This condition is also necessary, as is seen from

$$\sum_0^\infty \rho^\nu |\sin(\nu+\tfrac{1}{2})\theta| > \sum_0^\infty \rho^\nu \sin^2(\nu+\tfrac{1}{2})\theta = \frac{1}{2}\left\{\frac{1}{1-\rho} - \frac{(1-\rho)\cos\theta}{1-2\rho\cos\theta+\rho^2}\right\}$$

$$= \frac{1-(1+\rho^2)\cos\theta+\rho^2}{2(1-\rho)(1-2\rho\cos\theta+\rho^2)} = \frac{(1+\rho)^2(1-\cos\theta)}{2(1-\rho)\{(1-\rho)^2+2\rho(1-\cos\theta)\}}.$$

Thus the regularity condition here is the same as for Abel-summability.

The method corresponding to (1.6) is now

$$A_n(\rho_n, \theta_n) \equiv \sum_{\nu=0}^{n} u_\nu \rho_n^\nu \cos\nu\theta_n$$

$$= \sum_0^{n-1} s_\nu \rho_n^\nu \{\cos\nu\theta_n - \rho_n \cos(\nu+1)\theta_n\} + s_n \rho_n^n \cos n\theta_n; \tag{3.3}$$

the regularity conditions for this method are:

$$\rho_n \to 1, \qquad \theta_n \to 0 \qquad \text{as } n \to \infty, \tag{3.4}$$

and

$$\rho_n^n |\cos n\theta_n| + \sum_0^{n-1} \rho_n^\nu |\cos\nu\theta_n - \rho_n \cos(\nu+1)\theta_n| = O(1) \qquad \text{as } n \to \infty. \tag{3.5}$$

From $|\Re(x^\nu - x^{\nu+1})| \leqq |x^\nu - x^{\nu+1}|$ it is clear that the conditions (2.11) are sufficient for the regularity of the method (3.3). We shall prove that they are also necessary. We first show that $\rho_n^n = O(1)$ is a necessary condition. If $n|\theta_n| \leqq \frac{\pi}{4}$, then $\rho_n^n |\cos n\theta_n| \geqq \rho_n^n \frac{1}{\sqrt{2}}$. If $n|\theta_n| > \frac{\pi}{4}$, then define $\lambda = \lambda_n \geqq 1$ by

$$(2\lambda-1)\frac{\pi}{4} < n|\theta_n| \leqq (2\lambda+1)\frac{\pi}{4},$$

and define κ as the smallest integer for which

$$\kappa|\theta_n| \geqq (2\lambda-1)\frac{\pi}{4}.$$

Thus $(\kappa-1)|\theta_n| < (2\lambda-1)\frac{\pi}{4} < n|\theta_n|$, $\kappa < n$; and

$$\kappa \geqq (2\lambda-1)\frac{\pi}{4|\theta_n|} = \frac{(2\lambda+1)\pi}{4|\theta_n|}\cdot\frac{2\lambda-1}{2\lambda+1} \geqq \frac{2\lambda-1}{2\lambda+1}n \geqq \frac{n}{3}. \tag{3.6}$$

Furthermore

$$\sum_0^{n-1} \rho_n^\nu \,|\cos \nu\theta_n - \rho_n \cos(\nu+1)\theta_n|$$

$$\geqq \Big|\sum_0^{\kappa-1} (\rho_n^\nu \cos \nu\theta_n - \rho_n^{\nu+1} \cos(\nu+1)\theta_n)\Big| = |1 - \rho_n^\kappa \cos \kappa\theta_n|,$$

and

$$(2\lambda - 1)\frac{\pi}{4} \leqq \kappa\,|\theta_n| < (2\lambda - 1)\frac{\pi}{4} + |\theta_n|,$$

hence

$$\kappa\,|\theta_n| = (2\lambda - 1)\frac{\pi}{4} + \epsilon_n, \qquad 0 \leqq \epsilon_n < |\theta_n|,$$

$$\cos \kappa\,|\theta_n| \geqq \sin\left(\frac{\pi}{4} - |\theta_n|\right).$$

This and (3.6) proves that (3.5) implies $\rho_n^n = O(1)$. Now using (3.2) we see that (3.5) is equivalent to

$$\theta_n \sum_0^{n-1} \rho_n^\nu \,|\sin(\nu + \tfrac{1}{2})\theta_n| = O(1) \qquad \text{as } n \to \infty.$$

Hence we must have

$$\theta_n \sum_0^{n-1} \rho_n^\nu \sin^2(\nu + \tfrac{1}{2})\theta_n = O(1) \qquad \text{as } n \to \infty. \tag{3.7}$$

But

$$\sum_0^{n-1} \rho^\nu \sin^2(\nu + \tfrac{1}{2})\theta = \sum_0^{n-1} \rho^\nu(1 - \cos(2\nu+1)\theta) = \frac{1-\rho^n}{1-\rho} - \frac{(1-\rho)\cos\theta}{1+\rho^2 - 2\rho\cos 2\theta}$$

$$+ \frac{\rho^n[\cos(2n+1)\theta - \rho\cos(2n-1)\theta]}{1+\rho^2-2\rho\cos 2\theta}$$

$$= \frac{1-\rho^n}{1-\rho} - \frac{(1-\rho)\cos\theta}{(1-\rho)^2+4\rho\sin^2\theta} + \frac{\rho^n[\cos(2n+1)\theta - \cos(2n-1)\theta]}{(1-\rho)^2+4\rho\sin^2\theta}$$

$$+ \frac{\rho^n(1-\rho)\cos(2n-1)\theta}{(1-\rho)^2+4\rho\sin^2\theta}.$$

Now

$$\frac{|1-\rho_n|\,|\cos\theta_n|}{(1-\rho_n)^2+4\rho_n\sin^2\theta_n} \leqq \frac{|1-\rho_n|}{4\rho_n^{\frac{1}{2}}\,|1-\rho_n|\,|\sin\theta_n|} = O\left(\frac{1}{\theta_n}\right),$$

$$\frac{|\cos(2n+1)\theta_n - \cos(2n-1)\theta_n|}{(1-\rho_n)^2+4\rho_n\sin^2\theta_n} \leqq \frac{2\,|\sin\theta_n|}{4\rho_n\sin^2\theta_n} = O\left(\frac{1}{\theta_n}\right),$$

and we find that (3.7) implies $\theta_n \dfrac{1-\rho_n^n}{1-\rho_n} = O(1)$. We have thus proved:

THEOREM 2. *Necessary and sufficient conditions for the regularity of the transform* (3.3) *are conditions* (2.11). *In the special case* $\rho_n \equiv 1$ *these conditions reduce to* $\theta_n = O(1/n)$ *as* $n \to \infty$. The corresponding transform $A_n(1, \theta_n) = \sum_0^n u_\nu \cos \nu\theta_n$ was first introduced by Rogosinski [5, 6], in connection with trigonometric series.

4. Finally let

$$\varphi_n(x) = \frac{\sin nx}{nx}, \qquad n = 0, 1, 2, \cdots, (\varphi_0 \equiv 1).$$

Now (1.2) with $x \to 0$ is Lebesgue's summability; it is not regular, as (1.4) is not satisfied. The associated transform (1.6) becomes

$$A_n(x_n) = \sum_0^n u_\nu \frac{\sin \nu x_n}{\nu x_n}, \qquad n = 0, 1, 2, \cdots. \tag{4.1}$$

The necessary and sufficient conditions for regularity are (from (1.8) and (1.9) for real x_n):

$$x_n \to 0$$

and

$$\sum_{\nu=0}^{n} \left| \frac{\sin \nu x_n}{\nu x_n} - \frac{\sin (\nu + 1)\, x_n}{(\nu + 1) x_n} \right| = O(1) \qquad \text{as } n \to \infty.$$

We restrict our discussion to real positive $x_n = \theta_n \downarrow 0$; we shall prove that as in Theorem 2, $n\theta_n = O(1)$ is the necessary and sufficient condition for regularity. Elementary calculus shows that $\sin \theta/\theta$ is monotonic in the successive intervals $t_{\nu-1} \leqq \theta \leqq t_\nu$, $\nu = 1, 2, 3, \cdots$, where $t_0 = 0$ and $t_1, t_2, t_3, \cdots$ are the positive roots of the equation

$$\sin \theta = \theta \cos \theta; \tag{4.2}$$

also

$$t_\nu = \nu\pi + \alpha_\nu, \quad \text{where} \quad 0 < \alpha_\nu < \frac{\pi}{2},$$

and substitution into (4.2) yields easily

$$\frac{\pi}{2} - \alpha_\nu < \frac{1}{2\nu},$$

in particular $\alpha_\nu \to \pi/2$. Subdividing in

$$\sum_0^n \left| \frac{\sin \nu\theta_n}{\nu\theta_n} - \frac{\sin (\nu + 1)\theta_n}{(\nu + 1)\theta_n} \right|$$

the summation into parts, in each of which the differences have constant sign we find now that $n\theta_n = O(1)$ is the necessary and sufficient condition for uniform boundedness. Thus we have proved:

THEOREM 3. *The transform* (4.1) *with* $x_n = \theta_n \downarrow 0$ *is regular if and only if* $n\theta_n = O(1)$ *as* $n \to \infty$.

5. In this section we establish a lemma for later application. We restrict ourselves to real positive sequences $\{x_n\}$. We have seen already (§2) that the class

$$x_n^n = O(1) \qquad \text{as } n \to \infty, \tag{5.1}$$

is characterized by

$$\varliminf_{n\to\infty} n(1 - x_n) > -\infty; \tag{5.2}$$

and the smaller class

$$x_n^n = o(1) \qquad \text{as } n \to \infty, \tag{5.3}$$

is characterized by

$$n(1 - x_n) \to +\infty \qquad \text{as } n \to \infty. \tag{5.4}$$

We now prove the

LEMMA. *Let* $0 < x_n < 1$ *for all large* n; *then any one of the three conditions*:

$$x_n^n \log \frac{1}{1 - x_n} = O(1), \tag{5.5}$$

$$x_n^n = O\left(\frac{1}{\log n}\right), \qquad \text{as } n \to \infty, \tag{5.6}$$

$$\varliminf_{n\to\infty} \{n(1 - x_n) - \log\log n\} > -\infty, \tag{5.7}$$

implies the two others.

First assume (5.5); then

$$k_1 x_n^{-n} + \log(1 - x_n) > 0 \qquad \text{for all large } n. \tag{5.8}$$

Let

$$g(x) = k_1 x^{-n} + \log(1 - x), \qquad 0 < x < 1,$$

then

$$g'(x) = -nk_1 x^{-n-1} - \frac{1}{1 - x} < 0,$$

hence $g(x) \downarrow -\infty$ as $x \uparrow 1$. From (5.8): $g(x_n) > 0$ for all large n. We introduce for a constant k

$$x_n(k) \equiv 1 - \frac{1}{n}(\log\log n - k);$$

(2.10) yields

$$(5.9)\quad (x_n(k))^n = e^{n \log x_n(k)} = \exp\{(k - \log\log n)(1 + o(1))\} = \frac{e^k}{\log n}(1 + o(1))$$

as $n \to \infty$. Hence $g(x_n(k)) < 0$ for $e^{-k} < 1/k_1$ and all large n. Thus

$$(5.10)\qquad x_n < x_n(k) \text{ for } k > \log k_1\,, \text{ and for all large } n,$$

which is (5.7); moreover (5.9) and (5.10) yield (5.6). We now assume (5.6); thus

$$x_n^n < k_2/\log n \text{ for all large } n,$$

or

$$x_n^{-n} > \log n/k_2\,, \quad \text{and} \quad e^{n(x_n^{-n}-1)} > x_n^{-n} > \log n/k_2 \text{ for } n > k_3\,.$$

Hence

$$n(x_n^{-n} - 1) > \log\log n - \log k_2\,,$$

or

$$(5.11)\qquad x_n < \frac{1}{1 + \frac{1}{n}(\log\log n - \log k_2)} < 1 - \frac{1}{n}(\log\log n - \log k_2),$$

which gives (5.7). From here we get easily

$$(5.12)\qquad \log\frac{1}{1 - x_n} = \log n + O(1),$$

which together with (5.6) yields (5.5). Finally assume (5.7); then (5.11) holds, and the inequality

$$1 - x < e^{-x} \quad \text{for} \quad 0 < x < 1$$

gives $x_n^n < k_2/\log n$, which is (5.6). This and (5.12) finally give (5.5). This proves the lemma.

6. We now apply the summability methods introduced in §§2, 3, 4, to Fourier series. Let $f(\theta)$ be an integrable function of period 2π with the Fourier series:

$$(6.1)\qquad f(\theta) \sim \tfrac{1}{2}a_0 + \sum_1^\infty (a_\nu \cos\nu\theta + b_\nu \sin\nu\theta),$$

so that

$$b_0 = 0, \qquad a_\nu + ib_\nu = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta)e^{i\nu\theta}\,d\theta, \qquad \nu = 0, 1, 2, \cdots.$$

Applying (2.3) to the series (6.1) yields

$$(6.2)\qquad \begin{aligned} A_n(x_n, f) &\equiv \tfrac{1}{2}a_0 + \sum_{\nu=1}^{n} x_n^\nu(a_\nu \cos\nu\theta + b_\nu \sin\nu\theta) \\ &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(\theta + t)\left(\frac{1}{2} + \sum_{\nu=1}^{n} x_n^\nu \cos\nu t\right) dt \\ &= \frac{1}{\pi}\int_0^\pi \{f(\theta + t) + f(\theta - t)\}\, T_n(x_n, t)\, dt, \end{aligned}$$

where

$$T_n(x_n, t) = \frac{1}{2} + \sum_1^n x_n^\nu \cos \nu t.$$

Let

$$H(x, \theta) = \tfrac{1}{2}a_0 + \sum_1^\infty x^\nu(a_\nu \cos \nu\theta + b_\nu \sin \nu\theta),$$

$$S_n(x, \theta) = \tfrac{1}{2}a_0 + \sum_1^n x^\nu(a_\nu \cos \nu\theta + b_\nu \sin \nu\theta),$$

then $A_n(x_n, f) = S_n(x_n, \theta)$.

(6.2) is an integral-transform of the function $f(\theta)$ into a sequence A_n by the kernel $T_n(x_n, t)$. Such a transform is called regular, if $A_n \to \frac{1}{2}\{f(\theta) + f(-\theta)\}$ at every point of continuity. On defining [cf. Szász [8], chapter 3]

$$K(x, t) = \frac{2}{\pi} T_n(x_n, t) \quad \text{for } n \leqq x < n + 1, \quad \text{and} \quad 0 \leqq t \leqq \pi,$$

$$\text{and} \quad K(x, t) = 0 \quad \text{for } x \geqq 0, t > \pi,$$

furthermore

$$\chi(t) = \tfrac{1}{2}\{f(\theta + t) + f(\theta - t)\},$$

we get for the transform (6.2) $\int_0^\infty K(x, t)\chi(t)\, dt$, a step function of x. The necessary and sufficient conditions for regularity of this transform are [cf. Agnew 1].

(6.3) $$\lim_{n\to\infty} \int_h^\pi T_n(x_n, t)g(t)\, dt = 0 \quad \text{for any integrable } g(t) \text{ and any } h > 0,$$

(6.4) $$\overline{\lim_{n\to\infty}} \int_0^\pi |\, T_n(x_n, t)\,|\, dt < \infty,$$

(6.5) $$\lim_{n\to\infty} \int_0^\pi T_n(x_n, t)\, dt = \frac{\pi}{2}.$$

It is of interest to consider real positive sequences x_n, for which $T_n(x_n, t)$ is non negative for all t. Evidently there exists an r_n such that $T_n(x_n, t) \geqq 0$ for all t, when $0 < x_n \leqq r_n$, whereas for any $\epsilon > 0$ $T_n(r_n + \epsilon, t)$ becomes < 0 for some t. I. Schur and G. Szegö [7] have proved that:

(6.6) $$r_n \uparrow 1, \qquad \frac{n}{\log n}(1 - r_n) \to 1 \qquad \text{as } n \to \infty,$$

(6.7) $$r_n > 1 - \frac{1}{n}\log 2n \quad \text{for } n \geqq 1.$$

For odd n r_n is the positive root of the equation

$$1 - r - 2r^{n+1} = 0. \tag{6.8}$$

It follows from (6.8) and (6.6) that

$$r_n^n = O(1 - r_n) = O\left(\frac{\log n}{n}\right) \quad \text{as } n \to \infty. \tag{6.9}$$

Thus (from §2) the transforms (2.3) and (2.6) are both regular for $x_n = r_n$. Moreover $A_n(r_n, f)$ is a regular transform; this follows from the statement:

THEOREM 4. *The function-to-sequence transform* (6.2) (*where* $0 < x_n < 1$) *is regular if and only if* $x_n \to 1$ *and if* (5.7) *holds.*

Evidently for any x_n

$$\int_0^\pi T_n(x_n, t)\, dt = \frac{\pi}{2},$$

hence we need only satisfy (6.3) and (6.4). We restrict ourselves to $0 < x_n < 1$, and make use of the formula [7, §1]:

$$\begin{aligned} &2T_n(x, t) \\ &= (1 - 2x \cos t + x^2)^{-1}\{1 - x^2 + 2x^{n+2} \cos nt - 2x^{n+1} \cos (n+1)t\}. \end{aligned} \tag{6.10}$$

The expression $\frac{2}{\pi}\int_0^\pi |\, T_n(x_n, t)\,|\, dt = l(x_n)$ may be called the n^{th} Lebesgue constant of our summability method. As

$$\int_0^\pi \frac{1 - x^2}{1 - 2x \cos t + x^2} dt = \pi,$$

it is clear that (6.4) is satisfied if and only if

$$x_n^n \int_0^\pi \frac{|\, x_n \cos nt - \cos (n+1)t\,|}{1 - 2x_n \cos t + x_n^2} dt = O(1) \qquad \text{as } n \to \infty. \tag{6.11}$$

But from the identity

$$x \cos nt - \cos (n+1)t = x\{\cos nt - \cos (n+1)t\} - (1 - x) \cos (n+1)t$$

it follows that (6.11) holds if and only if

$$x_n^n \int_0^\pi |\cos nt - \cos (n+1)t|\, (1 - 2x_n \cos t + x_n^2)^{-1} dt = O(1). \tag{6.12}$$

We now prove that the necessary and sufficient condition for (6.12) is (5.5). To show that (5.5) implies (6.12) write

$$\begin{aligned} &\int_0^\pi |\cos nt - \cos (n+1)t|\, (1 - 2x_n \cos t + x_n^2)^{-1} dt \\ &= 2\int_0^\pi \sin \frac{t}{2} \frac{|\sin (n + \frac{1}{2})t\,|\, dt}{(1 - x_n)^2 + 4x_n \sin^2 \frac{t}{2}} < \int_0^\pi \frac{t\, dt}{(1 - x_n)^2 + \frac{4}{\pi^2} x_n t^2} \\ &< \int_0^{1-x_n} \frac{t\, dt}{(1 - x_n)^2} + \frac{2}{x_n} \int_{1-x_n}^\pi t^{-1} dt = \frac{1}{2} + \frac{2}{x_n} \log \frac{\pi}{1 - x_n}, \end{aligned}$$

for $0 < x_n < 1$; this proves the first part of our statement. The converse is seen from the estimate:

$$\int_0^\pi \sin\frac{t}{2}\,\frac{|\sin(n+\frac{1}{2})t|\,dt}{(1-x_n)^2+4x_n\sin^2\frac{t}{2}} > \frac{1}{2\pi}\int_{1-x_n}^{\pi}\frac{t\,|\sin(n+\frac{1}{2})t|\,dt}{t^2} > k\log\frac{1}{1-x_n}.$$

We finally show that (6.3) is satisfied whenever $x_n \to 1$ (assuming that (5.5) holds). We have

$$x_n^n\int_h^\pi |x_n\cos nt-\cos(n+1)t|\,|g(t)|\,(1-2x_n\cos t+x_n^2)^{-1}\,dt$$
$$< \frac{x_n^n}{1-\cos h}\int_0^\pi |g(t)|\,dt = o(1)$$

as $n\to\infty$ (from (5.6)). Hence, using (6.10), (6.3) holds if and only if

$$(1-x_n^2)\int_h^\pi (1-2x_n\cos t+x_n^2)^{-1}g(t)\,dt\to 0 \qquad \text{as } n\to\infty,$$

and this is equivalent to $x_n\to 1$. From the lemma of §5 we finally get Theorem 4.

We now apply the summability method (2.6) to Fourier series. Introducing the notation

$$s_0=\tfrac{1}{2}a_0, \qquad s_n(f,\theta)=s_n=\tfrac{1}{2}a_0+\sum_1^n (a_\nu\cos\nu\theta+b_\nu\sin\nu\theta), \quad n\geqq 1,$$

and using the formula

$$s_n=\frac{1}{2\pi}\int_0^\pi \{f(\theta+t)+f(\theta-t)\}\,\frac{\sin(n+\frac{1}{2})t}{\sin\frac{1}{2}t}\,dt,$$

we get from (2.6) the function-to-sequence transform:

$$B_n(x_n,f)\equiv\frac{1-x_n}{2\pi}\int_0^\pi \{f(\theta+t)+f(\theta-t)\}U_n(x_n,t)\,dt, \tag{6.13}$$

where

$$U_n(x_n,t)=\frac{1}{\sin\frac{1}{2}t}\sum_1^n x_n^\nu\sin(\nu+\tfrac{1}{2})t.$$

For the regularity of the transform (6.13) we have the necessary and sufficient conditions:

$$\lim_{n\to\infty}(1-x_n)\int_h^\pi U_n(x_n,t)g(t)\,dt=0 \quad \text{for any } g(t)\in L, \tag{6.14}$$

and any given $h>0$,

$$\overline{\lim_{n\to\infty}}\,\frac{1-x_n}{\pi}\int_0^\pi |U_n(x_n,t)|\,dt<\infty, \tag{6.15}$$

$$\lim_{n\to\infty}\frac{1-x_n}{\pi}\int_0^\pi U_n(x_n,t)\,dt=1. \tag{6.16}$$

For $f(t) \equiv 1$ formula (6.13) yields

$$1 - x_n^{n+1} = \frac{1 - x_n}{\pi} \int_0^\pi U_n(x_n, t)\, dt,$$

and condition (6.16) becomes $x_n^n \to 0$ as $n \to \infty$; assume again $0 < x_n < 1$, then we must have $n(1 - x_n) \to +\infty$. But Fejér [2, p. 20] remarked that for any $c_0 > c_1 > \cdots > c_n > 0$, $\sum_0^n c_\nu \sin(\nu + \frac{1}{2})t > 0$ for $0 < t < 2\pi$; hence the kernel of the transform (6.13) is positive, and the condition (6.15) is satisfied. We finally discuss condition (6.14). From the harmonic development

$$\sum_0^\infty x^\nu \sin(\nu + \tfrac{1}{2})t = \frac{(1 + x)\sin \frac{1}{2}t}{1 - 2x\cos t + x^2}, \qquad 0 < x < 1,$$

we find easily the formula

$$U_n(x, t)\sin \tfrac{1}{2}t(1 - 2x\cos t + x^2)$$
$$= (1 + x)\sin \tfrac{1}{2}t + x^{n+2}\sin(n + \tfrac{1}{2})t - x^{n+1}\sin(n + \tfrac{3}{2})t.$$

Using this formula and the condition $x_n^n \to 0$, it follows immediately that (6.14) holds if and only if

$$(1 - x_n)\int_h^\pi (1 - 2x_n\cos t + x_n^2)^{-1} g(t)\, dt \to 0,$$

i.e. if $x_n \to 1$. Thus we have

THEOREM 5. *The necessary and sufficient conditions for the regularity of the transform* (6.13) (*where* $0 < x_n < 1$) *are*:

$$x_n \to 1 \quad \textit{and} \quad n(1 - x_n) \to +\infty, \qquad \textit{as } n \to \infty.$$

7. We give here a Tauberian type theorem for the method $A_n(x_n)$ of §2.

THEOREM 6. *Suppose* x_n *real or complex, and*

$$\overline{\lim_{n\to\infty}} \,|1 - x_n|\,\frac{|x_n|^{-n} - 1}{1 - |x_n|} = \lambda < 1, \tag{7.1}$$

then $A_n(x_n) \to s$ *implies* $s_n \to s$.

Note that (7.1) implies $\overline{\lim}_{n\to\infty} ||x_n|^{-n} - 1| \leqq \lambda$, and in particular $x_n \to 1$, and

$$\frac{1}{1 + \lambda} \leqq \varliminf_{n\to\infty} |x_n|^n \leqq \overline{\lim_{n\to\infty}} |x_n|^n \leqq \frac{1}{1 - \lambda}.$$

The theorem can be deduced from recent results of R. Rado [4]; we give a direct proof, following his device.

Suppose first that $s = 0$ and $s_n = O(1)$; then $0 \leqq \overline{\lim}_{n\to\infty} |s_n| = \delta < \infty$; we shall prove that the assumption $\delta > 0$ leads to a contradiction. To a given $\epsilon > 0$ choose $n = n(\epsilon)$ so that $|s_\nu| < \delta + \epsilon$ for $\nu > n$; then choose $m > n$ so that $|s_m| > \delta - \epsilon$. Now, using (2.3)

$$\delta - \epsilon < |s_m| = \left| A_m x_m^{-m} - \sum_0^{m-1} s_\nu x_m^{\nu-m}(1-x_m) \right|$$

$$< |A_m x_m^{-m}| + \left| \sum_0^{n} s_\nu x_m^{\nu-m}(1-x_m) \right| + (\delta+\epsilon)|1-x_m| \frac{|x_m|^{-m}-1}{1-|x_m|};$$

thus, if m is large enough

$$\delta - \epsilon < \epsilon + \epsilon + (\delta+\epsilon)(\lambda+\epsilon) < \delta\lambda + \epsilon(3+\delta+\epsilon),$$

where $\lambda < 1$. This is a contradiction for ϵ small enough.

We next assume $s = 0$ and $\overline{\lim}\,|s_n| = \infty$; choose $0 < \epsilon$ small and l large; denote by $m = m(l)$ the least n for which $|s_n| > l$. Then (2.3) yields

$$l < |s_m| \leqq |A_m x_m^{-m}| + l|1-x_m| \frac{|x_m|^{-m}-1}{1-|x_m|} < \epsilon + l(\lambda+\epsilon),$$

which again is impossible for small ϵ. Thus the theorem is proved for $s = 0$. Finally for $s \neq 0$ we apply our result to the sequence $s_n - s$, and its transform

$$\sum_{\nu=0}^{n-1} (s_\nu - s)\, x_n^\nu (1-x_n) + (s_n - s) x_n^n = A_n - s \to 0,$$

which yields $s_n \to s$; our theorem is proved.

FINAL NOTE. The argument of §6 can be used to prove that the transforms $A_n(x_n, f)$, $B_n(x_n, f)$ under the regularity conditions of Theorems 4 and 5 converge to $f(\theta)$ if only

$$\int_0^t |f(\theta+u) + f(\theta-u) - 2f(\theta)|\, du = o(t) \qquad \text{as } t \to 0,$$

thus for almost all θ.

UNIVERSITY OF CINCINNATI

LITERATURE

1. R. P. AGNEW, *Properties of generalized definitions of limit*, Bull. Am. Math. Soc., vol. 45 1939, pp. 690–730.
2. L. FEJÉR, *Trigonometrische Reihen und Potenzreihen mit mehrfach monotoner Koeffizientenfolge*, Trans. Am. Math. Soc., vol. 39, 1936, pp. 18–59.
3. O. PERRON, *Beitrag zur Theorie der divergenten Reihen*, Math. Zeitschrift, vol. 6, 1920, pp. 286–310.
4. 4. R. RADO, *Some elementary Tauberian Theorems* (1), Quarterly Jour. of Math., Oxford Series, vol. 9, 1938, pp. 274–282.
5. W. ROGOSINSKI, *Über die Abschnitte trigonometrischer Reihen*, Math. Annalen, vol. 95, 1926, pp. 110–134.
6. W. ROGOSINSKI, *Abschnittsverhalten bei trigononetrischen und insbesondere Fourierschen Reihen*, Math. Zeitschrift, vol. 41, 1936, pp. 75–136.
7. I. SCHUR UND G. SZEGÖ, *Über die Abschnitte einer im Einheitskreise beschraenkten Potenzreihe*, Sitzungsberichte der preussischen Akademie, 1923, pp. 545–560.
8. O. SZÁSZ, *Lectures on summability methods*, University of Cincinnati, 1936–1937.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

GENERALIZED SURFACES IN THE CALCULUS OF VARIATIONS

BY L. C. YOUNG

(Received September 30, 1941)

I. GENERALIZED LIPSCHITZIAN SURFACES

1. Introduction

This note deals with an extension of the idea of surface, which is similar to the writer's generalization[1] of the notion of curve in the Calculus of Variations. The primary object of such generalizations is to ensure, as far as possible, that every problem of the Calculus of Variations has at least one solution. Their need has been apparent ever since the illuminating remark made by Hilbert:[2] *Eine jede Aufgabe der Variationsrechnung besitzt eine Lösung sobald hinsichtlich der Natur der Grenzbedingungen geeignete einschränkende Annahmen erfüllt sind, und, nötigenfalls, der Begriff der Lösung eine sinngemässe Erweiterung erfährt.*

In the existence theorems for variational problems concerning an unknown curve, the customary restriction to regular problems can now be removed by using *generalized curves*, and applications of these notions have lately been made by McShane.[3]

Problems of minima of variational problems concerning surfaces are *essentially* far more complicated than those concerning curves. In this note, however, the additional complication appears only in the proof of certain theorems, not in their statement.

The definition of *generalized surface* given here, and the proof of the existence of a (generalized) solution, are framed for a problem of minimum originally stated in a class Σ of ordinary surfaces

$$z = z(x, y)$$

with a given boundary. In the existence theorem, the principal further hypothesis is the restriction to a Lipschitzian class of surfaces, by which we mean that the function $z(x, y)$ satisfies the condition

$$| z(x', y') - z(x'', y'') | \leqq K\rho,$$

where K is a constant depending only on the class Σ and where ρ is the distance of the arbitrary points (x', y') and (x'', y''). The restriction expressed by this Lipschitz condition is rendered necessary by the ordinary (as opposed to the

[1] Young [12], [13], [14].
[2] Hilbert [2] p. 184; cf. also Lebesgue [3] p. 372.
[3] McShane [4], [5], [6].

parametric) formulation of the problem, not by the fact that we have surfaces rather than curves.

It is perhaps superfluous to add that the results of the present note merely constitute a beginning. The Lipschitz condition, although needed in the general type of problem considered here, is inconvenient in the applications to classical problems such as that of Dirichlet or Plateau, where a Lipschitzian boundary function may correspond to one or more singularities of the solution. And in any case, the usefulness of existence theorems which are based on a generalization, devised for the purpose, of the notion of surface, is naturally dependent on a new extension of the necessary conditions for attained minima such as that studied by the writer and by McShane in the case of curves. These questions we reserve for a later date.

2. An elementary unattained minimum

We shall follow the example set by McShane in his treatment of generalized curves[4] and begin with a "heuristic discussion." For the purpose of this discussion, let us compare our variational problem with an elementary minimum problem which consists in finding the least value of the function

$$f(t) = t^4 - 8t,$$

the latter being supposed defined in the first instance for *rational* t only. Let us analyze briefly this elementary problem as treated in the Differential Calculus, the term "least value" being first reinterpreted in the obvious way.

We first define *convergence* of a sequence of rational numbers, so as to make every such simple function $f(t)$ *continuous*. More precisely, whenever $\{t_n\}$ is a convergent sequence of rational numbers so is $\{f(t_n)\}$. Such a definition is furnished by the classical convergence criterion of Cauchy; indeed the classical idea of convergence may fairly be said to have been devised primarily for this very purpose. Next we define irrational numbers as limits of rational ones, and we extend to them the definition of the function $f(t)$ by continuity. In so doing we need not refer to properties special to numbers, except those implied in the definition of convergent sequence.[5] Finally we apply the methods of elementary Analysis to show that in its extended range the given function attains its least value, and to determine the latter in the usual way.

Can we proceed similarly in the Calculus of Variations? The above comparison strongly suggests that we can, and this without restricting ourselves in

[4] McShane [4] p. 514.

[5] The writer has in mind a definition "by abstraction" according to the terminology of Weyl [10] p. 7; and this definition is perhaps free from the criticism which Weyl himself [11] has directed against irrationals. Two sequences have the property of *limiting equality* if they are subsequences of a same convergent sequence; a system of sequences between any two of which the above property holds defines a real number, which is termed rational when the system contains a sequence of repetitions, otherwise irrational. With this definition, only the property of continuity actually stated is needed to extend $f(t)$ to irrational values of t.

any way to regular problems as past writers have done, if only we can produce a suitable definition of convergence. Now it is quite true that the definition of convergence normally applied to surfaces S in Analysis, renders the functions $F(S)$ which occur in variational problems, discontinuous; this is so even in regular problems, when the functions are merely semi-continuous. But, as remarked by McShane, we need not look very far for a more suitable definition: it is sufficient to term the sequence of surfaces $\{S_n\}$ convergent if and only if, for every $F(S)$ of the Calculus of Variations, the sequence of numbers $\{F(S_n)\}$ converges. The functions $F(S)$ are then, automatically, continuous.

3. Connection with linear operations

In classical problems of the Calculus of Variations, the functions $F(S)$ have the form

$$L_S(f) = \int\int_S f(x, y, z, p, q)\, dx\, dy = \int\int_A \bar{f}(x, y)\, dx\, dy,$$

where the integrand f is a continuous function of five real variables and $\bar{f}$ is its value at the point x, y of the domain[6] of definition A when we substitute for the three variables z, p, q the function $z(x, y)$ and its gradient $p(x, y), q(x, y)$. We shall suppose for simplicity that the domain A is *convex*. Actually only minor verbal changes are needed to make the discussion applicable to domains of the most general kind.[6a]

Just as in the writer's work on rectifiable curves, it is possible to treat problems of minimum in which a number of such double integrals occur simultaneously, by allowing the functions f and F to assume values in a vector-space instead of only real values; we shall content ourselves with the case of functions taking real values, and leave the reader to make the modifications appropriate to vector-functions which are suggested by the writer's treatment for curves.[7]

With the above form for $F(S)$, the definition of convergence of surfaces suggested in the preceding paragraph leads us to study linear operations $L(f)$ which are defined for every continuous function f. The definition itself, for the sequence of surfaces $\{S_n\}$, coincides with what is known as *weak convergence* of the operations $L_{S_n}(f)$, that is to say the convergence for each f of the sequence of numbers defined by these operations on f.

We shall suppose the functions f defined in the whole space of the five variables x, y, z, p, q as we clearly may; but we shall see in the next paragraph that no genuine linear operations, and no genuine weakly convergent sequences of linear operations, can exist, which are not wholly independent of the values taken by the functions f outside some fixed sphere $x^2 + y^2 + z^2 + p^2 + q^2 \leqq k^2$. This

[6] By a *domain* we mean a closed set of positive plane measure whose boundary is of plane measure zero.

[6a] The principal of these changes consists in interpreting *distance* of two points to mean minimum length of a path in A joining them.

[7] Young [13].

requires on the one hand that the domain A must be interior to a fixed circle of the x, y plane, and on the other hand that p, q be bounded, i.e. that the class of surfaces considered be *Lipschitzian*.

4. Linear operations in a certain type of space

Among the spaces which have become of interest in recent years, the best known are the Banach spaces, of which a particular case is the space whose elements are the continuous functions defined in a closed n-dimensional sphere. The space which is relevant in this note is one whose elements are the continuous functions defined in the *whole* of a five-dimensional space; in this case, any element f can be regarded as specified by means of a sequence $\{f_n\}$, where f_n denotes the continuous function defined in a sphere of radius n and fixed center, which coincides in its sphere with thc corresponding values of f.

We formulate the results of this paragraph rather more generally, in a space whose elements f are specified by sequences $\{f_n\}$, where f_n is a point which lies in a Banach space, possibly variable with n. Writing $|f_n|$ for the norm, or magnitude, of f_n in its space, we denote by $Q_n(f)$ the greatest of the first n numbers $|f_k|$; in the special case which mainly concerns us $Q_n(f)$ is the maximum absolute value of the function in the sphere of radius n. We write further

$$Q(f) = \sum 2^{-n} \frac{Q_n(f)}{1 + Q_n(f)}, \tag{4.1}$$

and we define the distance of two elements f, g to be $Q(f - g)$. In addition to verifying the usual laws of a distance, we observe at once from the definition of $Q(f)$ and from the fact that $Q_n(f)$ increases with n, that

(4.2) *If* $Q_N(f) \leqq 2^{-N}$ *for some* N, *then* $Q(f) \leqq 2^{-(N-1)}$.

This being so, let us consider the linear functions $L(f)$ defined in our space of elements f. We shall suppose the space *complete* when the distance is defined as above, a condition that is satisfied by the space of the continuous functions of five unlimited real variables. We shall further restrict each $L(f)$ to be measurable B, and therefore continuous[8] in f. Moreover, the absolute value $|L(f)|$ being continuous also, an important result[9] enables us to assert that, for a sequence $\{L_m(f)\}$ of such functions, which is bounded for each f of a set of the second category, there exists a sphere of elements f in which the sequence is *uniformly* bounded.

Writing $f = f_0 + g$ where f_0 is the center of this sphere, we see that the sequence $\{L_m(g)\} = \{L_m(f) - L_m(f_0)\}$ is uniformly bounded in the appropriate sphere of elements g with the origin as center, and *a fortiori* in the set $[Q_N(g) \leqq 2^{-N}]$, which by (4.2) is a subset of this sphere if N is large. By homogeneity, it follows that there is a constant K such that for every f, and for all m,

---

[8] Banach [1] p. 23 theorem 4.

[9] Banach [1] p. 19 theorem 11.

$$| L_m(f) | \leqq K \cdot Q_N(f).$$

We have thus proved the following theorem.

(4.3) *If the sequence of the values of the linear functions $L_m(f)$ is bounded for each f of a set of the second category, then there exist constants K and N independent of f and m, such that*

$$| L_m(f) | \leqq K \cdot Q_N(f). \tag{4.4}$$

As an immediate corollary, obtained by taking $\{L_m(f)\}$ to consist of repetitions of the linear function $L(f)$ defined for all f, we get

(4.5) *A linear function $L(f)$ defined for all the sequences $f = \{f_n\}$ depends only on a finite number of the f_n.*

In fact, if the points f and g denote two such sequences having the same N first terms, (4.4) applied to the point $f - g$, for which we have $Q_N(f - g) = 0$, gives $L_m(f - g) = 0$ where $L_m(f)$ is now a repetition of $L(f)$, and so $L(f) = L(g)$.

A set of linear functions $L(f)$ will be termed *compact* if every sequence of linear functions belonging to the set contains a subsequence which converges weakly. The following result is again a corollary of theorem (4.3).

(4.6) *The linear functions $L(f)$ of a compact set depend only on a bounded number of the f_n.*

For otherwise, we could form a sequence $L_n(f)$ of functions of the set such that (4.4) were false when $m = N = n$. This sequence would have to be unbounded at some f and so could not be compact, contrary to the hypothesis.

5. Generalized surfaces

We have already associated the notion of convergence of surfaces with the expressions $F(S)$ rather than with a geometrical representation of the surfaces concerned. Since the notion of generalized surface is to be attached to that of convergence, it is the process of calculation of the relevant $F(S)$ that we shall really be generalizing, rather than the geometrical representation.

Now this process of calculation is mainly that of passing from an arbitrary function of five variables $f(x, y, z, p, q)$ which we supposed continuous, to a function of two variables $\bar{f}(x, y)$ which we rendered measurable in A, by substituting for the three variables z, p, q the corresponding functions of x, y on S. From the function $\bar{f}$ we then passed to $F(S)$ by integration. In generalizing this process, we place, for reasons sufficiently explained elsewhere,[10] the substitution of $z(x, y)$ for z on a wholly different footing to that of its gradient for p, q. We still substitute for z a function $z(x, y)$ of which we say that it defines the track of the generalized surface S^*, but instead of substituting similarly for p, q we form a certain average of the values assumed by f at the various p, q.

By an *average*, or more precisely a *linear average* or *linear mean*, of a continuous function $g(p, q)$ of the variables p, q, we designate a functional $M(g)$,

[10] Young [13] p. 231.

linear in g, non-negative for non-negative g, and which satisfies the condition $M(g) = 1$ when g is the constant unity.

We term *measurable* in x, y, an average which varies according to the values of the additional parameters x, y, if for every continuous g which depends on p, q only, the number $M(g)$ obtained by forming the average at x, y defines a measurable function of x, y. As shown by McShane,[11] this number then still defines a measurable function when g is a continuous function of all four variables x, y, p, q.

This being so, we are now in a position to state our definitions. A *generalized surface* is the system defined by the following pair of elements: (i) a *real function* $z(x, y)$, absolutely continuous in the sense of Tonelli;[12] (ii) an *average* M defined for points x, y of A and measurable in x, y, which operates on each continuous function of p, q, and which satisfies for almost all x, y of A the condition that for the two functions $g(p, q) = p$ and $g(p, q) = q$ *the averages* $M(p)$, $M(q)$ *are the components of the gradient of* $z(x, y)$. It should be added that we identify two generalized surfaces when they have identical tracks $z(x, y)$ and for almost all x, y identical averages M. In the sequel, we shall often assume, without saying so explicitly, that the average has been suitably modified at the points x, y of a set of plane measure zero.

The generalized surface S^* will be termed *Lipschitzian with the constant* K, if at each point x, y the average $M(g)$ of an arbitrary function $g(p, q)$ is wholly independent of the values taken by $g(p, q)$ outside the circle of radius K whose center is the origin of the p, q plane.

With the definition of a generalized surface S^*, we give also that of *integral over* S^* *of the continuous function* $f(x, y, z, p, q)$, i.e. the meaning to be attached to the symbols

$$F(S^*) = L_{S^*}(f) = \int\int_A \bar{f}(x, y)\, dx\, dy.$$

We do this by stipulating that $\bar{f}(x, y) = M(g)$ where $g = g(p, q)$ is the value at the point x, y of the function $g(x, y, p, q)$ obtained by substituting $z(x, y)$ for z in $f(x, y, z, p, q)$. The fact that M is measurable in x, y ensures that the function $\bar{f}$ is so too.

Let us observe that an ordinary surface S and the corresponding function $F(S)$ are special cases of the above definitions: the average M then consists for almost all x, y in substituting in the function concerned the gradient of $z(x, y)$ for p, q.

6. Convergence of generalized surfaces

We now extend to the case of generalized surfaces the definition of convergence proposed in §2, and at the same time avoid conflict with the definition which is usual in Analysis. We shall consider only *Lipschitzian* generalized sur-

[11] McShane [4] p. 517 Lemma 3.1; the proof is given in the case of a single parameter t, but extends without difficulty to two parameters x and y.

[12] Saks [8] p. 169.

faces, this restriction being necessary in view of (4.7), in order that the repetitions of a given generalized surface constitute a sequence which converges according to our definition. We say that the sequence of generalized surfaces $\{S_n^*\}$ *converges* if the sequence $\{L_{S_n^*}(f)\}$ converges weakly, i.e. converges for every continuous function $f(x, y, z, p, q)$.

Just as we distinguish between a generalized surface S^* and its track, we distinguish also between convergence of generalized surfaces and *convergence of their tracks*, the latter notion being defined to be uniform convergence of the corresponding functions $z(x, y)$, as in the case of the usual definition of convergence of surfaces, which we thus identify with, and shall refer to in the sequel as, convergence of the tracks of surfaces.

Let us observe further that our definition of convergent sequence of generalized surfaces contains no reference to a limit. We shall see in §7, that the space of Lipschitzian generalized surfaces is complete, so that this kind of convergence may eventually be identified with the more familiar notion of convergence to a limit. In the meantime let us state explicitly that the term *compact set* denotes a set in which all infinite subsets contain corresponding convergent sequences, and that we do not restrict these sequences to possess limits at the moment. In the writer's opinion, this interpretation of the notion of compactness is in any case preferable to the one usually adopted, since this notion then becomes an intrinsic property of a set instead of a property which depends partly on a set and partly on the space.

This being so, we have the following theorem of compactness.

(6.1) *A system of generalized surfaces is compact if and only if their tracks lie in a fixed sphere and their Lipschitz constants are bounded.*

The pair of conditions stated is clearly equivalent to the single condition which restricts the corresponding linear operations $L_{S^*}(f)$ to be wholly independent of the values of the continuous functions f outside a fixed sphere of the x, y, z, p, q space. The necessity of these conditions therefore follows from (4.6). To prove their sufficiency, we suppose them satisfied, so that the linear operations $L_{S^*}(f)$ may be regarded as defined in the *Banach space* consisting of all functions of x, y, z, p, q which are continuous in a fixed sphere; the conclusion then reduces to a corollary of a well known theorem.[13]

We now consider some consequences of the notion of convergence. If S^* is any generalized surface, and we write as usual

$$L_{S^*}(f) = \iint_A \bar{f}(x, y)\, dx\, dy,$$

for the corresponding linear operation, we define more generally, for any interval Δ of the x, y plane, the *linear operation additive in* Δ

$$L_{S^*}(f; \Delta) = \iint_{A \cdot \Delta} \bar{f}(x, y)\, dx\, dy.$$

[13] Banach [1] p. 123 theorem 3.

We remark that for any fixed function f the operations corresponding to a compact system of generalized surfaces are *equi-continuous* in Δ. In fact, denoting by $K(f)$ the maximum modulus of the function f in a fixed sphere of the x, y, z, p, q space outside which the values of f are irrelevant by the preceding theorem, we clearly have

$$|L_{S^*}(f;\Delta) - L_{S^*}(f;\Delta')| \leqq K(f)\cdot|\Delta - \Delta'|. \tag{6.2}$$

We shall now establish the following result.

(6.3) *If the generalized surfaces S_n^* converge, then the operations $L_{S_n^*}(f;\Delta)$ converge for each f and uniformly in Δ.*

It is sufficient by equi-continuity, to establish the convergence for each f and each Δ. We enclose the interval Δ in a concentric similar interval Δ' of slightly larger area, and we denote by g a continuous function of x, y, z, p, q whose absolute value is majorized by that of f, such that for all z, p, q we have $g = f$ when x, y lies in Δ, and $g = 0$ when x, y lies outside Δ'. The existence of a function g with these properties is clear.

This being so, it follows from (6.2) that the expressions

$$L_{S_n^*}(f;\Delta) = L_{S_n^*}(g;\Delta)$$

differ from the corresponding terms of the *convergent* sequence formed by the expressions

$$L_{S_n^*}(g) = L_{S_n^*}(g;\Delta'),$$

at most by the arbitrarily small amount

$$K(g)\cdot|\Delta' - \Delta| \leqq K(f)\cdot|\Delta' - \Delta|,$$

and hence that they too must converge by Cauchy's convergence test. This completes the proof.

(6.4) *If the generalized surfaces S_n^* converge, so do their tracks.*

To see this, we observe in the first place that the functions z_n defining these tracks have by (6.1) uniformly bounded gradient, since their gradients at any point are averages of bounded sets of values of p, q. These functions are therefore equi-continuous, so that to show that they converge uniformly we need only verify that they converge at an everywhere dense set or that they converge in some average sense in the x, y plane. The latter is the case by (6.3) applied to the function $f(x, y, z, p, q) = z$, which states that the double integrals over Δ of the functions $z_n(x, y)$ converge uniformly.

7. The completeness theorem

We say that the sequence $\{S_n^*\}$ has the *limit* S^* if the operations $L_{S_n^*}(f)$ have the limit $L_{S^*}(f)$ for each f. Evidently the existence of the limit implies the convergence of the sequence; we shall establish the converse.

(7.1) *A convergent sequence of generalized surfaces has a limit.*

We denote by $z(x, y)$ the limit of the corresponding tracks, and by $L(f;\Delta)$

that of the corresponding operations additive in Δ, the limits existing by (6.4) and (6.3). We denote further by $Q(f;\ \Delta)$ the maximum modulus of the function f when $z = z(x,\ y)$, when $x,\ y$ lies in $A \cdot \Delta$, and when $p,\ q$ is bounded in magnitude by the upper bound of the Lipschitz constants of our sequence, finite by (6.1).

Clearly the operation $L(f;\ \Delta)$ is linear in f, additive in Δ, and we have

$$| L(f;\ \Delta) | \leqq | \Delta | \cdot Q(f;\ \Delta), \tag{7.2}$$

as the limit of similar relations.

It follows from (7.2) that $L(f;\ \Delta)$ is the double integral over Δ of any network derivative $M(f;\ x,\ y)$, and that the latter exists almost everywhere, as the limit, for a suitable sequence of intervals Δ tending to the point $x,\ y$, of the expression

$$L(f;\ \Delta)/| \Delta | \tag{7.3}$$

when the function f is kept fixed.

For almost every $x,\ y$, this limit exists simultaneously for all functions f of a given enumerable everywhere dense set of functions. Since the functionals (7.3) are, by (7.2), equi-continuous in f for the various Δ, the derivative in question also exists simultaneously for all f at any such point $x,\ y$; and it thus represents, for almost any fixed $x,\ y$, a linear operation in f which, as we deduce at once from (7.2), is majorized by the maximum modulus, for the same system of $p,\ q$ as before, of the function f when $x,\ y$ and $z = z(x,\ y)$ are fixed. Since it has this majorant, the operation $M(f;\ x,\ y)$ depends on the values of f as function of $p,\ q$ only, the other variables being kept fixed.

This being so, let us write simply $M(f)$ for the operation just defined at almost every $x,\ y$, and let us complete its definition for the remaining points $x,\ y$ by choosing it for instance to mean the value of f for $p = q = 0$.

We verify at once that $M(f)$, at the point $x,\ y$ and $z = z(x,\ y)$, is a non-negative operation when f is non-negative for all $p,\ q$ at this point, and that for the function $f = 1$ we have $M(f) = 1$. Hence $M(f)$ is an average. Moreover we have already remarked that its double integral over any Δ coincides with $L(f;\ \Delta)$. In order to identify the latter with the operation $L_{S^*}(f;\ \Delta)$, and thus to complete the proof of (7.1), it is now sufficient to verify that a generalized surface S^* is defined by the track $z(x,\ y)$ and the average $M(f)$. This will be the case if the averages of the functions $f = p$ and $f = q$ are the components of the gradient of $z(x,\ y)$ at almost all points $x,\ y$ of the domain of definition A.

To establish this last result, it is sufficient to show that for these two functions f and for an arbitrary interval Δ contained in the domain A, the expression $L(f;\ \Delta)$ coincides with the double integral over Δ of the corresponding component of the gradient. Since the corresponding relations are true for the sequence of which L is the limit, it only remains to verify that the double integral over Δ of the gradients of the tracks of the generalized surfaces of our sequence tend to the double integral of the gradient of $z(x,\ y)$. This is evidently the case in view of the uniform convergence of these tracks.

8. Closure of the class of ordinary surfaces

From the theorem of completeness just established, it follows in particular that the limit of any convergent sequence of *ordinary* surfaces necessarily exists and represents a generalized surface whose Lipschitz constant is majorized by the upper bound of those occurring in the sequence. We now consider the converse question, whether *every* Lipschitzian generalized surface is expressible as the limit of a suitable sequence of ordinary surfaces. Our next theorem shows that the answer is in the affirmative, even with additional restrictions on this sequence.

(8.1) *The class of the generalized surfaces whose tracks have a fixed boundary and whose Lipschitz constants do not exceed N, is identical with the closure of the subclass consisting of ordinary surfaces.*

It is sufficient to prove that, given any generalized surface S^* with a Lipschitz constant not exceeding N, there exists a sequence of ordinary surfaces S_n with Lipschitz constants not exceeding N and boundary coinciding with that of the track of S^*, such that, for all continuous functions f of the variables x, y, z, p, q,

$$L_{S^*}(f) = \lim_n L_{S_n}(f). \tag{8.2}$$

Let us say that a generalized surface $S^{*\prime}$ is an ϵ-approximation to the generalized surface S^*, if there exist a positive function η of ϵ only, which tends to 0 with ϵ, such that

$$| L_{S^*}(f) - L_{S^{*\prime}}(f) | \leqq K\eta \cdot Q(f) + K \cdot \omega(f; \eta), \tag{8.3}$$

where K is some fixed constant, and where $Q(f)$ and $\omega(f; \eta)$ denote, for the function f, the maxima, in some fixed sphere of x, y, z, p, q, of the modulus and of the oscillation between two points distant less than η.

In order to establish (8.2), and so (8.1), it is evidently enough to show that, for any positive ϵ, there is an ϵ-approximation to the given generalized surface S^* provided by an ordinary surface subject to the conditions stated for S_n. We shall divide the proof into several stages which occupy §§9–13.

9. A lemma due to McShane[14]

In the sequel we frequently have to enlarge and fit together disconnected portions of surfaces, without unduly increasing their Lipschitz constants. In so doing, we make use of the following lemma.

(9.1) *Suppose that a function $z(x, y)$, originally given in the plane set E, satisfies, for all pairs of points of this set, the Lipschitz condition*

$$| z(x', y') - z(x'', y'') | \leqq K \cdot \rho, \tag{9.2}$$

where ρ denotes the distance of the pair of points, and where K is a fixed constant. Then there exists a function $z(x, y)$, defined in the whole plane and coinciding with

[14] McShane[7] p. 838, Theorem I.

the given function in the set E, such that (9.2) *continues to hold with this same value of K for all pairs of points of the plane.*

For a proof of this lemma, we refer the reader to McShane.[14] We shall deduce a slight refinement of some importance to us. We remark that the continuation whose existence is asserted is by no means unique, although there are certain points of the plane, those of the original set E in particular, at which all such continuations necessarily coincide. These points constitute a set $\underline{E}$ that we shall call *set of unicity* for the required continuation.

(9.3) *There exists a continuation for which, in addition, every point outside the the set of unicity is the center of a circle in which* (9.2) *is valid with some smaller constant in place of K.*

To see this, we first observe that, for any given point not in $\underline{E}$, the validity of (9.2) with two distinct values $c \pm \epsilon K$ for $z(x', y')$ when we take x', y' at this point and x'', y'' to lie in the set E, leads to the inequality

$$| z(x'', y'') - c | \leqq K \cdot (\rho - \epsilon),$$

from which it follows that the function with the constant value c in the circle of radius ϵ around the given point, and with the value of the original function $z(x, y)$ at each point of E, satisfies (9.2) in the set consisting of E together with this circle. Consequently, applying (9.1) to this function, we see that given any point not in $\underline{E}$ there exists a continuation which is constant in some neighborhood of this point.

This being so, we can cover the complement of $\underline{E}$ by a sequence of such neighborhoods. Denoting by $z_n(x, y)$ the continuation which has a constant value in the n^{th} neighborhood, the function

$$\sum 2^{-n} \cdot z_n(x, y)$$

constitutes a continuation for which in the n^{th} neighborhood (9.2) is valid with the constant K reduced to $(1 - 2^{-n}) \cdot K$. This proves our assertion (9.3).

The reduction of K is not in general possible in the set $\underline{E}$. In fact, we have the following result.

(9.4) *A point of $\underline{E}$ at a positive distance from the set E is always contained in a segment consisting of points of $\underline{E}$ for every pair of which the relation* (9.2) *is an equality.*

Moreover, at such a point, the gradient of $z(x, y)$, if it exists, has magnitude K.

We remark, in the first place, that at a point x', y' of the set of unicity the value $z(x', y')$ is the only one which satisfies (9.2) for every x'', y'' of E; for otherwise we could add this point to the set E with either of two values for $z(x', y')$ and use (9.1) to form two different continuations corresponding to these two values.

Hence, if P is a point of $\underline{E}$ distant $d > 0$ from the set E, then given any positive ϵ there exists a point x'', y'' of E such that, if ρ'' is its distance from P,

$$| z(x'', y'') - z(P) | > K \cdot (\rho'' - \epsilon);$$

and from this last relation together with (9.2) we deduce that if x', y' is the point on the segment joining P to x'', y'' at distance d from the point P,

$$| z(x', y') - z(P) | > K \cdot (d - \epsilon).$$

These relations are true for any continuation $z(x, y)$. Making ϵ tend to 0, we see by continuity that there exists a point Q independent of the continuation chosen such that

$$| z(Q) - z(P) | = K \cdot d,$$

the distance of the points P and Q being again equal to d. It is clear moreover that the values of $z(x', y')$ corresponding to the various continuations can only differ by at most K, and hence that the value $z(Q)$ is unique. Finally it is clear from our last equation that (9.2) becomes an equality for every pair of points of the segment PQ, and that this segment lies in the set of unicity $\underline{E}$.

It now only remains to prove the second part of (9.4).[15] Let R denote the point at distance h from P along a parallel to the x-axis, and let k denote the distance RQ. We write further φ and θ for the angles made by RQ and PQ with the x-axis, and m for the ratio $| z(P) - z(R) |/h$. We have then

$$h \cdot m = | z(P) - z(R) | \geqq | z(P) - z(Q) | - | z(R) - z(Q) |$$
$$\geqq K \cdot d - K \cdot k = K \cdot \{(h^2 + k^2 + 2hk \cos \varphi)^{\frac{1}{2}} - k\}.$$

Hence, dividing by h and making h tend to 0, we find, since k then tends to d and φ to θ, that the lower limit of m is not less than the value

$$K \cdot \lim \frac{(h^2 + k^2 + 2hk \cos \varphi)^{\frac{1}{2}} - k}{h} = K \cdot \cos \theta.$$

Hence the x-component of the gradient of $z(x, y)$ at P is not less than $K \cdot \cos \theta$, and similarly the y-component is not less than $K \cdot \sin \theta$. This completes the proof.

10. Preliminary reduction

We say that the generalized surface S^* coincides with its track at the point x, y or in the subset B of A if at this point, or almost everywhere in this subset B, the average $M(f)$ defining S^* coincides for every continuous f with the result of substituting for p, q in f the gradient of the track of S^*. We remark that this is certainly the case of any point x, y at which the gradient has the magnitude N of the Lipschitz constant of S^*.

At such a point the gradient has the form $N \cos \alpha$, $N \sin \alpha$, whence for the function $g(p, q) = N - p \cos \alpha - q \sin \alpha$, we have $M(g) = 0$; and since, in the circle of relevant p, q, this function is non-negative and vanishes only for $p, q = N \cos \alpha, N \sin \alpha$, it follows in the usual way that $M(g) = 0$ for every $g(p, q)$ which vanishes at this point; so that, for any g with the value c at this point, we have $M(g - c) = 0$ i.e. $M(g) = c$, as asserted.

[15] The writer owes the second part of this proof to Mr. J. H. Whiteman.

This being so, we denote by E the boundary of A. The function $z(x, y)$ which defines the track of S^* is a continuation of the type (9.1) for its boundary values in E, the constant K of (9.2) being replaced by N. We form also a second continuation $Z(x, y)$, possibly identical with $z(x, y)$, of the type (9.3).

In the set of unicity $\underline{E}$ we certainly have $z(x, y) = Z(x, y)$. In this set, moreover, since E is closed and of measure zero, almost all points have positive distances from E; and since the gradient exists almost everywhere, its magnitude is N by (9.4) for almost all points of $\underline{E}$. Thus S^* coincides with its track in $\underline{E}$.

We write P, Q for the gradient of $Z(x, y)$ where it exists, and choose $P = Q = 0$ otherwise. We write further

$$p', q' = \epsilon P + (1 - \epsilon)p, \qquad \epsilon Q + (1 - \epsilon)q$$

and for any function $g(p, q)$ we define

$$M'(g) = M(g') \quad \text{where} \quad g'(p, q) = g(p', q').$$

With these definitions, the function

$$z'(x, y) = \epsilon Z(x, y) + (1 - \epsilon)z(x, y)$$

has almost everywhere the gradient $M'(p)$, $M'(q)$. Finally we denote by A' a subset of $A - \underline{E}$ consisting of a finite sum of squares in each of which $Z(x, y)$ satisfies a Lipschitz condition with a constant smaller than N. Since every point of $A - \underline{E}$ is the center of such a square by (9.3) we can choose A' so that its measure differs by less than ϵ from that of $A - \underline{E}$.

In A' the magnitude of P, Q has an upper bound N'' less than N, so that for any p, q of magnitude not exceeding N the magnitude of p', q' cannot exceed the value $N' = \epsilon N'' + (1 - \epsilon)N$ which is less than N. Hence for the points x, y of A', the average $M'(g)$ is wholly independent of the values taken by $g(p, q)$ at the points p, q of magnitude exceeding N', where N' is less than N.

This being so, we now define the generalized surface $S^{*\prime}$ in the following manner: the *track* of $S^{*\prime}$ is $z'(x, y)$ in A; the *average* is $M'(f)$ at each point x, y of A'; while in $A - A'$ the surface $S^{*\prime}$ coincides with its track. This generalized surface has the property that *its portion in A' has a Lipschitz constant less than N, while the remaining portion is an ordinary surface with the Lipschitz constant N.* We shall see that it constitutes an ϵ-approximation to S^*.

To this effect we observe that the relevant pairs of values p', q' and p, q have the difference $\epsilon(P - p)$, $\epsilon(Q - q)$ whose magnitude cannot exceed the number $\eta = 2\epsilon N$; therefore in A' we have

$$| M'(f) - M(f) | = | M(f' - f) | \leqq \text{Max} | f(p', q') - f(p, q) | \leqq \omega(f; \eta).$$

In $\underline{E}$ on the other hand the two generalized surfaces coincide with a same track, while the remainder of A, the set $A - A' - \underline{E}$, has small measure. From these facts, we easily obtain a relation of the type (8.3), which shows that $S^{*\prime}$ is an ϵ-approximation to S^*, as asserted.

Since this new generalized surface has moreover the same boundary as the

original one, we may substitute it to the latter for the proof of (8.1). In doing so we may at the same time replace the domain A by the finite sum of squares A', and since the latter can be approximated by a sum of squares without common points it will be sufficient to deal with the case in which A' consists of a single square.

11. Approximation by averages restricted to a finite set

We now make a second reduction which has an interest independent of (8.1). We consider a triangulation of the p, q plane by equilateral triangles of side ϵ, and denote by p_m , q_m, where $m = 1, 2, \cdots, n$, the vertices whose magnitude does not exceed N. An average $M(g)$ will be said to be restricted to the p_m , q_m, if it is of the form

$$\sum a_m g(p_m , q_m) \quad \text{where} \quad a_m \geqq 0 \quad \text{for each } m \text{ and where} \quad \sum a_m = 1.$$

We shall obtain the following result.

(11.1) *A generalized surface S^* with Lipschitz constant less than N has an ϵ-approximation $S^{*\prime}$ with the same track and with a Lipschitz constant less than N, such that, at each point x, y, the average M' defining $S^{*\prime}$ is restricted to the p_m , q_m.*

We shall suppose ϵ so small that the triangles with the vertices p_m , q_m enclose all values of p, q which are relevant to the averages $M(f)$ of the definition of S^*. We write in the form

$$(p, q) = \sum b_m \cdot (p_m , q_m)$$

where only the vertices of the triangle containing p, q have non-zero coefficients, the barycentric representation of the point p, q in its triangle or edge of triangle. We define further, for any $g(p, q)$,

$$M'(g) = M(g'), \quad \text{where} \quad g'(p, q) = \sum b_m \cdot g(p_m , q_m);$$

here M is the average occurring in the definition of S^* at any point x, y and this average is defined for the function g' since the latter is continuous, as we easily verify.

Now in the particular case where g is one of the functions p, q or the constant 1, we see at once that $g' = g$, and so $M(g) = M'(g)$. Hence, remembering that the coefficients of the barycentric expression are non-negative, M' is an average with which we can associate the same track as that of S^* to form an $S^{*\prime}$. Evidently $S^{*\prime}$ has a Lipschitz constant less than N, so that it only remains to prove that $S^{*\prime}$ is an ϵ-approximation to S^*. This however follows at once from the relations

$$| M'(g) - M(g) | = | M(g - g') | \leqq \sum b_m \cdot \omega(g; \epsilon) = \omega(g; \epsilon).$$

It is convenient to prove also the following more special result.

(11.2) *If the generalized surface S^* has a piece-wise constant gradient, then the approximation $S^{*\prime}$ satisfying the requirements made in* (9.1) *can be made to fulfill the additional condition that the finite sum $\sum a_m \cdot g(p_m , q_m)$ which constitutes the*

average $M'(g)$ at the point x, y shall have coefficients a_m which are piece-wise constant in x, y.

To see this, we divide any interval of constancy of the gradient of the track S^* into similar intervals of diameter less than ϵ', and denote by $\underline{a}_m$ the function of x, y which is the mean value of a_m in the interval of the division which contains the point x, y. As ϵ' tends to 0, the $\underline{a}_m$ tend boundedly to the a_m almost everywhere and the convergence is uniform in a subset whose measure is as close as we please to that of the set of definition A. From this it follows that the generalized surface derived from $S^{*\prime}$ by replacing the a_m by $\underline{a}_m$ fulfills all our requirements.

12. Reduction to a plane track

The principal part of this section does not concern generalized surfaces. We prove the following theorem.

(12.1) *The track of a Lipschitzian surface with constant less than N is expressible as the limit of a certain track $z_\epsilon(x, y)$ in such a manner that its gradient is almost everywhere the limit of that of $z_\epsilon(x, y)$, where the track defined by $z_\epsilon(x, y)$ satisfies the following additional conditions:*

(i) *its Lipschitz constant is less than N,*

(ii) *outside measure ϵ its gradient is piecewise constant.*

Moreover when these conditions are satisfied, any S^* with a Lipschitz constant less than N and with the original track, has a corresponding ϵ-approximation with Lipschitz constant less than N whose track is $z_\epsilon(x, y)$. Finally the theorem remains true if we restrict throughout the track $z_\epsilon(x, y)$ to have the same boundary as the original track.

The final remark concerning the boundary is an immediate corollary of the rest of the theorem in view of (9.1), if we apply it in the first place to a domain interior to A whose distance from the boundary of A is positive, and then continue the track obtained suitably to agree with the original track on the boundary of A. We may therefore ignore this part of the assertion, and since we can continue the given track by (9.1) into a square containing A we may suppose that A is such a square.

This being so, let $z(x, y)$ define the original track in A and let $\varphi(x, y)$ be the vector function consisting of the gradient of z, which exists almost everywhere in A. There exists a vector function $\Phi(x, y)$ piecewise constant in A; we shall suppose it constant in the squares R of side a, such that

$$|\varphi - \Phi| < \epsilon^{10} \quad \text{except possibly in measure} < \epsilon^{10}. \tag{12.2}$$

Let us denote by B the set of those squares R if any for which a subset of R of measure not less than $a^2 \cdot \epsilon^8$ is occupied by points for which the relation $|\varphi - \Phi| < \epsilon^{10}$ is false. Clearly this relation is then false in a subset of B of measure not less than $\epsilon^8 |B|$, so that by (12.2) we must have $|B| < \epsilon^2$.

Consider now any square R not in B. We have then

(12.3) $\quad |\varphi - \Phi| < \epsilon^{10}$ except possibly in measure $< a^2 \cdot \epsilon^8$ of R.

We denote by E the linear set contained in the projection of R on the y-axis which consists of the values of y for each of which the relation $|\varphi - \Phi| < \epsilon^{10}$ is false for a corresponding set of x of linear measure not less than $\epsilon^4 a$. It follows from (12.3) that the linear measure of E is less than $\epsilon^4 a$. The same being true when the axes of x and y are interchanged, it follows that we can determine a series of parallels to the two axes, forming a grating over R, in such a way that R is divided into rectangles of sides less than $\epsilon^4 a$ by this grating and that on any segment of this grating in R we have

(12.4) $\quad |\varphi - \Phi| < \epsilon^{10}$ except possibly in linear measure $< \epsilon^4 a$.

Now any point x, y of R can be joined to the center of R by a path C consisting of two portions of segments of the grating and two small segments, the lengths of the latter being less than $\frac{1}{2}\epsilon^4 a$. It follows from (12.4) that on the path C we have

(12.5) $\quad |\varphi - \Phi| < \epsilon^{10}$ except possibly in linear measure $< 3\epsilon^4 a$.

This being so, let $Z(x, y)$ denote the linear function defined in the interior of R to agree with $z(x, y)$ at the center of R and to have the constant gradient Φ. The difference of the functions z and Z at the point x, y cannot exceed in magnitude the integral on C of that of their gradients. Since these gradients are clearly less than N (if ϵ is sufficiently small), it follows from (12.5) that

$$|z(x, y) - Z(x, y)| \leqq \epsilon^3 \cdot a, \tag{12.6}$$

provided that ϵ is sufficiently small.

For every R not in B, we now denote by R' the concentric square of side $(1 - 2\epsilon^2) \cdot a$; we denote further by B' the set of the points of the squares R', and by B'' the set of the points of the boundaries of the corresponding squares R. In the set $B + B' + B''$ we define the function $z_\epsilon(x, y)$ by stipulating that

$$z_\epsilon(x, y) = z(x, y) \text{ in } B + B'', \qquad z_\epsilon(x, y) = Z(x, y) \text{ in } B'. \tag{12.7}$$

This function satisfies in $B + B' + B''$ the Lipschitz condition (9.2) with a constant K less than N; for we need only verify this when the two points x', y' and x'', y'' lie respectively in B' and in $B + B''$. But since the distance ρ of the points then exceeds $\epsilon^2 \cdot a$, we have

$$\begin{aligned} |z_\epsilon(x', y') - z_\epsilon(x'', y'')| &= |Z(x', y') - z(x'', y'')| \\ &\leqq \epsilon^3 \cdot a + |z(x', y') - z(x'', y'')| \\ &\leqq \epsilon \cdot \rho + K \cdot \rho < N \cdot \rho, \end{aligned}$$

which is the required verification.

This being so, we can continue the function $z_\epsilon(x, y)$ by (9.1) so that it still satisfies the same Lipschitz condition. Hence, observing that any point of A

outside B' is distant at most $\epsilon^2 \cdot a$ from one of coincidence of the functions $z_\epsilon(x, y)$ and $z(x, y)$, we see that the two functions differ by at most $2N \cdot \epsilon^2 a < \epsilon \cdot a$ in $A - B'$ and so by (12.6) throughout A.

Now in B', except possibly in a subset of measure ϵ^{10}, the gradient of $z_\epsilon(x, y)$ is piecewise constant and differs by less than ϵ^{10} from the gradient of $z(x, y)$. To show that $z_\epsilon(x, y)$ fulfils the requirements of our theorem, it is sufficient to verify that the complement of B' in A has measure less than $\epsilon - \epsilon^{10}$. This is the case since

$$| A - B' | = | B | + \{1 - (1 - 2\epsilon^2)\}^2 \cdot | A - B | \leqq \epsilon^2 + 4\epsilon^2 | A |.$$

Now finally let $z_\epsilon(x, y)$ denote any track fulfilling the requirements of the theorem, and let S^* denote a generalized surface with the original track.

We denote by η the lower bound of the numbers such that outside measure η at most

(12.8) the difference of the functions $z(x, y)$, $z_\epsilon(x, y)$ and that of their gradients are both of magnitude at most η.

From our hypotheses, it follows that η tends to 0 with ϵ. We write further A' for the set of x, y in which (12.8) holds, and p', q' for the result of translating p, q by the difference of these gradients. Denoting by M the average defining S^* at any point, we define, for any continuous $g(p, q)$, the average

$$M'(g) = M(g'), \quad \text{where} \quad g'(p, q) = g(p', q')$$

to be at any point of the set A' the average defining $S^{*\prime}$, and we complete the latter by making it coincide with its track in $A - A'$.

The surface $S^{*\prime}$ with the track $z_\epsilon(x, y)$ and the average thus defined, is easily verified to be defined consistently and to furnish the required ϵ-approximation.

13. Completion of the proof of (8.1)

The various reductions made possible by the results of the preceding sections show that it is now sufficient to establish (8.1) in the case of a generalized surface S^* defined for the points x, y of a square A, possessing a Lipschitz constant less than N, possessing in view of (12.1)—where we may now choose A to be a square of constancy of the gradient—a plane track; and finally possessing an average at x, y restricted to at most n vectors p_m, q_m, this average being of the form $\sum a_m g(p_m, q_m)$ where the coefficients a_m may be supposed piece-wise constant, and so, by subdivision, absolutely constant, in view of (11.2). With these additional hypotheses, which now cause no real loss of generality, we shall prove our theorem by an induction with respect to the number n.

It will be sufficient to prove that *S^* has an ϵ-approximation consisting of a generalized surface $S^{*\prime}$ formed of a finite number of portions in each of which the corresponding average $M'(g)$ is restricted to at most $n - 1$ of the p_m, q_m.*

Moreover we may clearly suppose that A is a triangle instead of a square. Further, we need not insist that the boundary of $S^{*\prime}$ is to coincide with that of S^*, since this can always be arranged by means of a subsequent modification of

$S^{*\prime}$ near its boundary, a modification which is made possible by (9.1) and which is, by now, sufficiently clear.

This being so, we may suppose by trivial transformations that the track of S^* is in the x, y plane, and that for a particular value of m, the value $m = 1$ say, we have $q_m = 0$ and $p_m > 0$. We suppose also that the coefficients a_m in the average M are all different from 0, for if this were not the case there would be nothing to prove.

We write k for an index to be summed from $k = 1$ to $k = n - 1$, and we define numbers a_0 and p_0 by the equations

$$a_0 = \sum a_{k+1}\,, \qquad a_0 p_0 = \sum a_{k+1} p_{k+1}\,.$$

These numbers then satisfy the equations

$$a_0 + a_1 = 1, \qquad a_0 p_0 + a_1 p_1 = 0, \qquad a_0 > 0, \qquad a_1 > 0.$$

We write further $b_k = a_{k+1}/a_0$ and observe that

$$\sum b_k p_{k+1} = p_0\,, \qquad \sum b_k q_{k+1} = 0$$

We now divide the plane into strips parallel to the y-axis whose widths are alternately ϵa_0 and ϵa_1, and we define a function of x only, $z'(x)$, linear in each strip and taking on the edges separating the strips alternately the values 0 and

$$-\delta = \epsilon a_0 p_0 = -\epsilon a_1 p_1\,,$$

from which it follows that the gradient of this function in the strips is alternately p_0 and p_1.

We define the generalized surface $S^{*\prime}$ to have in the triangle A the track $z'(x, y) = z'(x)$ and the average M' given by

$$M'(g) = \sum b_k g(p_{k+1}\,, q_{k+1})$$

in the strips where the gradient is p_0, while at the remaining points the surface coincides with its track i.e. $M'(g) = g(p_1\,, 0)$. Evidently these conditions are compatible with the definition of our generalized surfaces, since we find that the functions $g = p$ and $g = q$ have the averages p_0 and 0 in the strips where the gradient is p_0. Since the average M' is restricted in each strip to less than n vectors p_m, q_m it only remains to show that $S^{*\prime}$ is an ϵ-approximation to S^*.

We observe that on the segment parallel to the x-axis intercepted by a pair of consecutive strips at the height y, the integrals of the two averages $M(f)$ and $M'(f)$ are, for any continuous $f(x, y, z, p, q)$, products of the length of this segment by the expressions of the form

$$\sum a_m f(x, y, 0, p_m, q_m) \quad \text{and} \quad a_1 f(x', y, z', p_1, q_1) + a_0 \sum b_k f(x'', y, z'', p_{k+1}, q_{k+1})$$

where x, x', x'' are intermediate values of x, on account of the mean value theorem of the integral calculus, and where z', z'' are values of z which are small with ϵ. Since $a_m = a_0 b_{m-1}$, these expressions differ from one another by at most $\sum a_m \cdot \omega(f; \eta)$ i.e. by $\omega(f; \eta)$ where η tends to 0 with ϵ.

From this it easily follows that the corresponding double integrals over the triangle A differ by at most $K \cdot \omega(f; \eta)$ and so, that $S^{*\prime}$ is an ϵ-approximation of S^*, as required. This completes the proof.

14. Existence of an attained minimum. Equivalence of generalized and ordinary problem

Let us suppose that we are given the problem of the minimum of one of our functions $F(S)$ in the class of Lipschitzian ordinary surfaces with assigned boundaries. Beside this "ordinary" problem let us consider the "generalized problem" of the minimum of the function $F(S^*)$ in the corresponding class of generalized surfaces S^*.

There exists a sequence of ordinary surfaces $\{S_n\}$ of the class considered, for which the numbers $F(S_n)$ tend to the minimum of the first problem; but, by (8.1), this is also the case in the second problem. The two minima are thus identical, i.e.

(14.1) *The ordinary problem is equivalent to the generalized problem.*

We have established this result in the case in which the surfaces considered are Lipschitzian with unrestricted constants. The argument holds equally when their Lipschitz constants are restricted to be less than or at most equal to some fixed number K. In the latter case we can assert more.

(14.2) *In the class of surfaces with an assigned boundary whose Lipschitz constants do not exceed a fixed number K, the minimum of $F(S)$ is the same for generalized surfaces as for ordinary surfaces and there exists a generalized surface in the class for which this minimum is attained.*

For by (6.1) the minimizing sequence $\{S_n\}$ is compact. Denoting by S^* the limit of a subsequence we evidently have $F(S^*) = \operatorname{Lim} F(S_n)$ and this completes the proof.

Let us remark that in the case of a *regular* problem, it is easily seen that the solution S^* must coincide with its track, i.e. reduces to an ordinary surface. This does not however include the whole of what is known of the existence theory for the regular case, since Tonelli[16] obtained a semi-continuity theorem independent of the Lipschitz condition that we have had to assume here. Nevertheless, we shall attempt to show in a subsequent note that even for regular problems the present methods have distinct advantages.

CAPETOWN, SOUTH AFRICA

REFERENCES

[1] BANACH, S. *Théorie des opérations linéaires*, Monografje Matematyczne (Warszawa, 1932).

[2] HILBERT, D. *Über das Dirichletsche Prinzip*, Jahresber. der Deutschen Math. Vereinigung, Vol. VIII (1900) pp. 184–188.

[16] Tonelli [9] p. 342.

[3] LEBESGUE, H. *Sur le problème de Dirichlet*, Rendiconti del Circolo Matematico di Palermo, Vol. 24 (1907) pp. 371–402.

[4] MCSHANE, E. J. *Generalized curves*, Duke Math. Jour., Vol. 6 (1940) pp. 513–536.

[5] MCSHANE, E. J. *Necessary conditions in the generalized-curve problem of the Calculus of Variations*, Ibidem, Vol. 7 (1940) pp. 1–27.

[6] MCSHANE, E. J. *Existence Theorems for Bolza problems in the Calculus of Variations*, Ibidem, Vol. 7 (1940) pp. 28–61.

[7] MCSHANE, E. J. *Extension of range of functions*, Bull. of the Am. Math. Soc., Vol. 40 (1934) pp. 837–842.

[8] SAKS, S. *Theory of the Integral*, Monografje Matematyczne, Vol. VII (Warszawa, 1937).

[9] TONELLI, L. *Sur la semi-continuité des intégrales doubles du Calcul des Variations*, Acta Math., Vol. 53 (1929) pp. 325–346.

[10] WEYL, H. *Die Idee der Riemannschen Fläche*, Mathematische Vorlesungen an der Universität Göttingen, Vol. V (Teubner, Leipzig 1913).

[11] WEYL, H. *Über die neue Grundlagenkrise der Mathematik*, Math. Zeits., Vol. 10 (1921) pp. 39–79.

[12] YOUNG, L. C. *On approximation by polygons in the Calculus of Variations*, Proc. of the Royal Soc. (A), Vol. 141 (1933) pp. 325–341.

[13] YOUNG, L. C. *Generalized Curves and the existence of an attained absolute minimum in the Calculus of Variations*, Comptes Rendus de la Société des Sciences et des Lettres de Varsovie, Classe III, Vol. 30 (1937) pp. 212–234.

[14] YOUNG, L. C. *Necessary conditions in the Calculus of Variations*, Acta Math., Vol. 69 (1938), pp. 239–258.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

FREE LATTICES II

BY PHILIP M. WHITMAN

(Received July 15, 1941)

In the previous paper under this title [14] the author discussed the basic internal structure of free lattices. We assume here a knowledge of the notation used in that paper.[1] In the present paper the previous results are slightly extended and then applied to obtain further properties of free lattices. We study the automorphisms of free lattices (§2), their sublattices (§3), and their order topology (§4), where we find that in the terminology of Birkhoff [3] free lattices are continuous (perhaps vacuously) but not complete.

The author is indebted to Prof. Garrett Birkhoff for inspiration and advice in the preparation of this paper.

1. Introduction

The main definition of [14] will bear repetition: *the free lattice generated by* $x_1, x_2, \cdots, x_n$ *is a lattice generated by them in which there are no laws of equality except those derivable from the postulates for a lattice.* In [14] this lattice was denoted by F_n, since to within isomorphism it depends only on n and not on the symbols x_i. It now seems advisable to replace F_n by $FL(n)$, and also in some cases to specify the generators: $FL(x_1, x_2, \cdots, x_n)$, or $FL(X)$ where $X = \{x_1, x_2, \cdots, x_n\}$. One may also consider free modular or distributive lattices, denoted $FM(n)$ and $FD(n)$ and defined as are free lattices except that the modular or distributive law ([3] pp. 34, 74) is adjoined to the lattice postulates; cf. Dedekind [7]; Skolem [12]; Birkhoff [1] pp. 450–2, 463–4, [3] pp. 4, 49, 84–85, [4] p. 451; Church [5]. The existence of free algebras in general is shown in Birkhoff [2] p. 441.

The main results of the previous paper [14] may be restated (with slight extensions) as follows, starting with the corollary of §2 and the theorem of §3.

THEOREM 1: *In* $FL(n)$, $x_i \leqq x_j$ *if and only if* $i = j$; *recursively,*

(A) $\prod a_i \leqq x_j$ *if and only if some* $a_i \leqq x_j$;
(B) $x_i \leqq \sum b_j$ *if and only if* $x_i \leqq$ *some* b_j;
(C) $\sum a_i \leqq b$ *if and only if every* $a_i \leqq b$;
(D) $a \leqq \prod b_j$ *if and only if* $a \leqq$ *every* b_j;
(E) $\prod a_i \leqq \sum b_j$ *if and only if* $\prod a_i \leqq$ *some* b_j *or some* $a_i \leqq \sum b_j$.

THEOREM 2: *Of all the polynomials equal in* $FL(n)$ *to a given polynomial, there is one of shortest length, unique to within commutativity and associativity.*

This is taken as the canonical form. We write $a \equiv b$ for polynomials a and b

[1] Or in [3], plus the notation $L(a)$ = length of a for the number of generators appearing in the polynomial a, counting repetitions.

if one is obtainable from the other by commutativity and associativity. As a criterion for canonicity, we can weaken the conditions of Corollary (18) of [14] slightly:

COROLLARY 1.1: *If* $\sum a_j$ *is*

$$(\prod_k a_{1k}) \cup (\prod_k a_{2k}) \cup \cdots \cup (\prod_k a_{mk}) \cup x_{i_{m+1}} \cup \cdots \cup x_{i_u}$$

then $\sum a_j$ *is canonical if and only if* (A) *for all* j *and* p, $a_{jp} \not\leqq \sum a_k$, *and* (B) $a_j \leqq a_k$ *implies* $j = k$, *and* (C) *every* a_j *is canonical. Dually for* $\prod a_j$. *A generator by itself is canonical.*

PROOF: To prove this equivalent to the corollary cited, we need only to show that the failure of (b) in that corollary implies the failure of one of the conditions of Corollary 1.1. Thus suppose that in $FL(n)$, $a_i \leqq \sum_{k \neq i} a_k$.

CASE 1: $a_i \equiv x_j$. Then by Theorem 1 B, $x_j \leqq a_k$ for some k contrary to (B).

CASE 2: $a_i \equiv \prod_j a_{ij}$. Then by Theorem 1 E, either $\prod_j a_{ij} \leqq a_k$ for some k contrary to (B), or

$$a_{ij} \leqq \sum_{k \neq i} a_k \leqq \sum_{\text{all } k} a_k$$

for some j, contrary to (A).

COROLLARY 1.2: *If* $\sum a_i = \prod b_j$ *in* $FL(n)$ *and* $\sum a_i$ *is canonical, then* $\prod b_j = b_k$ *for some* k.

PROOF: This is a companion to Corollary (19) of [14] and is proved similarly.

LEMMA 1.1: *In* $FL(n)$ *for* n *finite, given the polynomial* a *and index* p *then either* $x_p \leqq a$ *or* $a \leqq \sum_{i \neq p} x_i$, *the two being mutually exclusive.*

PROOF: This is obvious for $L(a) = 1$; we proceed by induction on $L(a)$.

CASE 1: $a \equiv \prod a_i$. If $a \not\leqq \sum_{i \neq p} x_i$, then for all j, $a_j \not\leqq \sum_{i \neq p} x_i$, so by induction $x_p \leqq a_j$ for all j; thus $x_p \leqq \prod a_j = a$ as desired.

CASE 2: $a \equiv \sum a_i$. If $a \not\leqq \sum_{i \neq p} x_i$, then for some k, $a_k \not\leqq \sum_{i \neq p} x_i$, so by induction $x_p \leqq a_k \leqq a$ and thus $x_p \leqq a$. Thus at least one of the alternatives holds; they are mutually exclusive since $x_p \leqq a \leqq \sum_{i \neq p} x_i$ would imply $x_p \leqq \sum_{i \neq p} x_i$ and then by Theorem 1, $x_p \leqq x_i$ for some i not p which is impossible.

This is a special case of a "splitting" of a lattice [13], a subject which will be discussed in another paper.[2] An immediate consequence of this lemma is Corollary 3 of §4 of [14], for suppose that for all p, $a \not\leqq \sum_{i \neq p} x_i$; then by Lemma 1.1 we have $x_p \leqq a$ for all p, so $a \geqq \sum_{i=1}^{n} x_i$; since the latter is the unit element of $FL(n)$, we have $a = \sum_{i=1}^{n} x_i$, as asserted. The dual of this corollary says that the "points" or "atoms" ([3], p. 10) of $FL(n)$ are the $\prod_{i \neq p} x_i$. We can now strengthen Theorem 3 of [14], not only by extending it farther down in the lattice (which seems relatively uninteresting) but also in the content of its statements about the relation of other elements of the lattice to certain special elements.

THEOREM 3: *In* $FL(n)$ *for* n *finite, if* a *is given then either* $a = \sum_{i=1}^{n} x_i$ *or*

[2] Soon to be submitted for publication.

$a = \sum_{i \neq p} x_i$ for some p or else for some distinct p and q, $a = b_{pq}$ or $a \leqq c_{pq}$, where

$$b_{pq} \equiv (\sum_{i \neq p} x_i) \cap (\sum_{i \neq q} x_i)$$

$$c_{pq} \equiv (\sum_{i \neq p,q} x_i) \cup \sum_{r \neq p,q} [(\sum_{i \neq p} x_i) \cap (\sum_{i \neq q} x_i) \cap (\sum_{i \neq r} x_i)].$$

Moreover, if $a < \sum_{i \neq p} x_i$ then $a \leqq b_{pq}$ for the same p and some q, and likewise if $a < b_{pq}$ then $a \leqq c_{pq}$.

PROOF: In view of Corollary 3 of §4 of [14], discussed above, it suffices to prove the last sentence of the theorem; the rest will then follow. Hence we suppose $a < \sum_{i \neq p} x_i$. By Lemma 1.1, if q is given then either $a \leqq \sum_{i \neq q} x_i$ or $x_q \leqq a$. If the latter held for all q different from p then $\sum_{q \neq p} x_q \leqq a \leqq \sum_{i \neq p} x_i$ and so $a = \sum_{i \neq p} x_i$ contrary to hypothesis. Hence for some q, $a \leqq \sum_{i \neq q} x_i$ as well as $a \leqq \sum_{i \neq p} x_i$, and so $a \leqq b_{pq}$ for some q other than p, as desired. That $a < b_{pq}$ implies $a \leqq c_{pq}$ is proved by similar but more elaborate methods; since we do not use this result we omit the details.

For reference we note certain immediate consequences of Theorem 1; n is not required finite.

COROLLARY 1.3: *In $FL(n)$,* (A) *$\prod a_i \leqq \sum y_j$ if and only if for some i, $a_i \leqq \sum y_j$, and dually, where the y_j are a non-void subset of the x_i;* (B) *if $\sum a_i = x_j$ then $a_k = x_j$ for some k.*

PROOF: (A) If $a_i \leqq \sum y_j$ for no i, then by Theorem 1 E we would have $\prod a_i \leqq y_j$ for some j, and so for some i, $a_i \leqq y_j \leqq \sum y_k$. (B) By Theorem 1 B, $x_j \leqq a_k$ for some k; $a_i \leqq x_j$ for all i, so $a_k = x_j$ by (2) of [14].

2. Automorphisms of $FL(n)$

In some cases an algebra is the free algebra ([3] p. 4; [2] p. 441) for more than one set of generators;[3] for instance the free group generated by x and y is the same as that generated by x and $x^{-1}y$. We assert that this is not the case with free lattices. First however we must clarify the meaning of "$\{w_i\}$ is a set of generators." In [14] we said that a set $\{w_i\}$ generates a lattice L if the polynomials in the w_i exhaust L. Now if L consists say of all polynomials in some elements x_i, do we demand that every polynomial a in the x_i should be formally identical with some polynomial in the w_i after substituting the values of the w_i in terms of the x_i? No, this seems excessive; rather we demand that some polynomial in the w_i should *equal* a. (As a matter of fact the next theorem holds whichever interpretation is used.) Likewise, as regards automorphisms we consider only their effect on classes of equal elements, not how they may permute polynomials within such a class.

In this section we consider two sets of generators to be the same if their elements are equal by pairs.

DEFINITION 2.1: A set of generators of a lattice L is *redundant* if some member of the set can be omitted without affecting the property that the set generates L.

[3] The author is not aware of any precise reference in print, but this is well known. It is mentioned by J. Dyer-Bennet, thesis, *A free algebra with three operations*, unpublished, Harvard University, 1940.

THEOREM 4: *$FL(n)$ is the free lattice on a unique set X of generators; any set of generators must contain X; and X is the only irredundant set of generators.*

PROOF: First we observe that $FL(n)$ could not be the free lattice on a redundant set of generators, for if it were then by Definition 2.1 one of the generators would equal a polynomial in the others, which is not true in a free lattice. Now let $FL(n)$ be $FL(x_1, \cdots, x_n)$; that is, it consists of all lattice polynomials in the x_i, subject to the laws stated in Theorem 1; and suppose that it is also a lattice generated by $w_1, \cdots, w_m$. We assert that given k, then $w_i = x_k$ for some i. For suppose not: for some k and all i, $w_i \neq x_k$; then we may proceed by induction on $L(a)$ where a is any polynomial in the w_i. Suppose $a \neq x_k$ for all a with $L(a) < h$. Then suppose $a = x_k$ and $L(a) = h$; $a \equiv \sum a_i$ or dually. By Corollary 1.3 B, $a_i = x_k$ for some i, since a is indirectly a polynomial in the x_i, contrary to hypothesis of induction. Thus by induction we would have that no polynomial in the w_i equals x_k so the w_i could not be a set of generators. Hence $w_i = x_k$ for some i, given k, proving the second clause of the theorem. If moreover $\{w_i\}$ is irredundant then $\{w_i\}$ must coincide with $\{x_i\}$ to within equality, which proves the last clause. Then the first part of the proof proves the first clause. Q.E.D.

This states only that a different irredundant set of generators cannot generate *all* of $FL(n)$; as we shall see in the next section, they may generate a sublattice isomorphic to it. Another consequence is this: any automorphism of a lattice is by definition determined by the images of the generators, so by Theorem 3 an automorphism of $FL(n)$ must merely permute the x_i among themselves. On the other hand, by the symmetry of the postulates any such permutation determines an automorphism. Hence

COROLLARY 2.1: *The group of automorphisms of $FL(n)$ is the symmetric group on n symbols.*

3. Sublattices of $FL(n)$

We observe first that in free lattices the relations between polynomials in $x_1, \cdots, x_k$ are independent of the presence or absence of other generators:

LEMMA 3.1: *If $f(x_1, \cdots, x_k) \leqq g(x_1, \cdots, x_k)$ for polynomials f and g in $FL(n)$, then this relation is also true in $FL(k)$, and conversely.*

PROOF: By Theorem 1 the conditions in the two cases are the same.

LEMMA 3.2: *If $\{x_1, \cdots, x_r\}$, $\{y_1, \cdots, y_s\}$, and $\{z_1, \cdots, z_t\}$ are disjoint subsets of the generators of $FL(n)$, and*

$$f(x_1, \cdots, x_r, y_1, \cdots, y_s) \leqq g(x_1, \cdots, x_r, z_1, \cdots, z_t)$$

for polynomials f and g in $FL(n)$, then also

$$f(x_1, \cdots, x_r, 1, \cdots, 1) \leqq g(x_1, \cdots, x_r, 0, \cdots, 0)$$

in $FL(n)$.[4]

[4] 0 and 1 stand for the zero and unit elements of the lattice. If not present, they may for the purposes of this lemma be temporarily adjoined; cf. MacNeille [9] p. 443. They will nearly always be absorbed in reducing the new polynomials to simpler form.

PROOF: This is trivial for $L(f) + L(g) = 2$. Denoting the results of the substitutions by bars, we make an induction on $L(f) + L(g)$. If $g \equiv \prod g_i$, then by Theorem 1 $f \leqq g_i$ for all i, so by induction $\bar{f} \leqq \bar{g}_i$ for all i, and hence $\bar{f} \leqq \prod \bar{g}_i = \bar{g}$. If $g \equiv \sum g_i$ and $f \equiv \prod f_i$, then $f \leqq g_i$ for some i, or $f_i \leqq g$ for some i, so $\bar{f} \leqq \bar{g}_i$ or $\bar{f}_i \leqq \bar{g}$, and thus $\bar{f} \leqq \bar{g}$. The other cases are similar. Q.E.D.

Let us now consider the sublattices of $FL(x_1, \cdots, x_n)$ for n fixed. In particular, we inquire which of these sublattices are again free lattices on suitable generators. Since a sublattice is often most easily designated by listing its generators, we shall find it convenient to make the following

DEFINITION 3.1: *The set* $\{u_i\}$ *is free* if the sublattice generated by it is the free lattice with these generators.

The question then becomes: what conditions on the u_i will insure that they form a free set? In Theorem 5 it appears that the answer is simply that with respect to their joins and meets they shall behave as do the x_i. The situation in free lattices is thus unlike that in free groups, where every subgroup is again free.[5]

THEOREM 5: *A subset* $U = \{u_i\}$ *of the elements of* $FL(n)$ *is free if and only if* $u_j \leqq \sum_{i \in S} u_i$ *and its dual (where* S *is a finite set of indices) each imply* $j \in S$.

PROOF: We wish to show these conditions necessary and sufficient that in the given lattice $FL(n) \equiv FL(X)$, the sublattice generated by U is $FL(U)$, or equivalently that this sublattice is isomorphic to $FL(Y) \equiv FL(y_1, \cdots, y_k)$ where k is the cardinal power of the set U. Thus we wish to show that the conditions of the theorem are necessary and sufficient in order that for all polynomials f and g,

$$f(u_{i_1}, \cdots, u_{i_j}) \leqq g(u_{i'_1}, \cdots, u_{i'_j})$$

in $FL(X)$ if and only if

$$f(y_{i_1}, \cdots, y_{i_j}) \leqq g(y_{i'_1}, \cdots, y_{i'_j})$$

in $FL(Y)$, or (more briefly) in order that $f(U) \leqq g(U)$ in $FL(X)$ if and only if $f(Y) \leqq g(Y)$ in $FL(Y)$. The necessity of the given conditions is apparent from the fact that by Theorem 1 the y_i behave in this manner. For sufficiency we assume the conditions of the theorem and prove the isomorphism by induction on $L(f(Y)) + L(g(Y))$. If $L(f(Y)) + L(g(Y)) = 2$ then $L(f(Y)) = L(g(Y)) = 1$. By hypothesis $u_i \leqq u_j$ if and only if $i = j$, which is true if and only if $y_i \leqq y_j$, as desired. Proceeding by induction,

CASE 1: $f(Y) \equiv y_j$. Now $y_j \leqq g(Y)$ implies $u_j \leqq g(U)$ because the lattice generated by U is a homomorphic image of $FL(Y)$ by [2], Theorem 9 on p. 441. Conversely if the latter holds then the former does, for if not then by Lemmas 1.1 and 3.1 applied to the finite set R consisting of the y_i which appear in either $f(Y)$ or $g(Y)$,

---

[5] Nielsen [10], Schreier [11]. Levi [8] gives a new proof. The falsity for free lattices is shown by taking any two-element sublattice.

$$g(Y) \leqq \sum_{i \in R-j} y_i$$

in $FL(y_i\,;\, i \in R)$ and so also in $FL(Y)$. But then

$$u_j \leqq g(U) \leqq \sum_{i \in R-j} u_i$$

contrary to hypothesis.

CASE 2: $g(Y) \equiv y_j$. Dual of case 1.

CASE 3: $f(Y) \equiv \sum f_i(Y)$. Then $\sum f_i(U) \leqq g(U)$ implies by Theorem 1 that $f_i(U) \leqq g(U)$ for all i, so by induction $f_i(Y) \leqq g(Y)$ for all i, whence $f(Y) \leqq g(Y)$, while the converse follows as at the beginning of Case 1.

CASE 4: $g(Y) \equiv \prod g_i(Y)$. Dual of case 3.

CASE 5: $f(Y) \equiv \prod f_i(Y)$ and $g(Y) \equiv \sum g_i(Y)$. Then $f(U) \leqq g(U)$ implies by Theorem 1 that $f_i(U) \leqq g(U)$ for some i or else $f(U) \leqq g_i(U)$ for some i, and by induction we again get $f(Y) \leqq g(Y)$, and conversely as above. Q.E.D.

In particular we can verify that in $FL(3)$ these conditions are satisfied by $x_1 \cup (x_2 \cap x_3)$, $x_2 \cup (x_1 \cap x_3)$, and $x_3 \cup (x_1 \cap x_2)$, or still better,

LEMMA 3.3: *In $FL(3)$ the following are a free set*:

$$u_1 \equiv [x_1 \cap (x_2 \cup x_3)] \cup [x_2 \cap (x_1 \cup x_3)]$$

$$u_2 \equiv [x_1 \cap (x_2 \cup x_3)] \cup [x_3 \cap (x_1 \cup x_2)]$$

$$u_3 \equiv [x_1 \cup (x_2 \cap x_3)] \cap [x_2 \cup (x_1 \cap x_3)]$$

$$u_4 \equiv [x_1 \cup (x_2 \cap x_3)] \cap [x_3 \cup (x_1 \cap x_2)].$$

The proof is a matter of straightforward verification of the conditions of Theorem 5, using Theorem 1. We give one example: $u_4 \nleqq u_1 \cup u_2 \cup u_3$. For otherwise one of the following must hold: (*i*) $u_4 \leqq u_1$; (*ii*) $u_4 \leqq u_2$; (*iii*) $u_4 \leqq u_3$; (*iv*) $x_1 \cup (x_2 \cap x_3) \leqq u_1 \cup u_2 \cup u_3$; (*v*) $x_3 \cup (x_1 \cap x_2) \leqq u_1 \cup u_2 \cup u_3$, by Theorem 1 E. But if (*i*) held, then either $u_4 \leqq x_1 \cap (x_2 \cup x_3)$ or $u_4 \leqq x_2 \cap (x_1 \cup x_3)$ or $x_1 \cup (x_2 \cap x_3) \leqq u_1$ or $x_3 \cup (x_1 \cap x_2) \leqq u_1$, which are contradicted respectively by the facts that $u_4 \nleqq x_1$, $u_4 \nleqq x_2$, $x_1 \nleqq u_1$, and $x_3 \nleqq u_1$, so (*i*) is impossible. Similarly so is (*ii*). If (*iii*) held, then $u_4 \leqq x_2 \cup (x_1 \cap x_3)$; thus $u_4 \leqq x_2$ or $u_4 \leqq x_1 \cap x_3$ or $x_1 \cup (x_2 \cap x_3) \leqq x_2 \cup (x_1 \cap x_3)$ or $x_3 \cup (x_1 \cap x_2) \leqq x_2 \cup (x_1 \cap x_3)$. These are respectively contradicted by the facts that $u_4 \nleqq x_2$, $u_4 \nleqq x_1 \cap x_3$, $x_1 \nleqq x_2 \cup (x_1 \cap x_3)$, and $x_3 \nleqq x_2 \cup (x_1 \cap x_3)$, so (*iii*) is false. (*iv*) and (*v*) are false since $x_1 \nleqq u_1 \cup u_2 \cup u_3$ and $x_3 \nleqq u_1 \cup u_2 \cup u_3$. Thus the typical condition chosen holds; the others are similar. Q.E.D.

THEOREM 6: *$FL(3)$ has $FL(n)$ as a sublattice, for any finite or countable n.*

PROOF: By Lemma 3.3, $FL(3)$ has a free set $\{u_1, u_2, u_3, u_4\}$. We proceed by induction on the number of elements in a free set $\{u_i\}$. Suppose $\{u_1, u_2, \cdots, u_k\}$ is a free set of elements of $FL(3)$. By Lemma 3.1, $\{u_{k-2}, u_{k-1}, u_k\}$ is free, so by Lemma 3.3 there is a free set $\{v_1, v_2, v_3, v_4\}$ of polynomials in u_{k-2}, u_{k-1}, and u_k. Then

$$\{u_1, u_2, \cdots, u_{k-3}, v_1, v_2, v_3, v_4\}$$

is free, for we may readily verify the conditions of Theorem 5:

CONDITION 1: $u_1 \not\leqq u_2 \cup u_3 \cup \cdots \cup u_{k-3} \cup v_1 \cup v_2 \cup v_3 \cup v_4$.

For otherwise

$$u_1 \leqq u_2 \cup \cdots \cup u_{k-3} \cup v_1 \cup v_2 \cup v_3 \cup v_4 \leqq u_2 \cup \cdots \cup u_k$$

contrary to hypothesis that $\{u_1, \cdots, u_k\}$ was free.

CONDITION 2: $v_4 \not\leqq u_1 \cup u_2 \cup \cdots \cup u_{k-3} \cup v_1 \cup v_2 \cup v_3$.

For otherwise (since v_4 is a polynomial in u_{k-2}, u_{k-1}, and u_k alone) by Lemma 3.2, $v_4 \leqq v_1 \cup v_2 \cup v_3$ contrary to the choice of the v_i.

The other conditions follow by symmetry and duality. Thus the conditions for induction are fulfilled and we obtain a countable set of u_i. This whole set together is free, for the conditions of Theorem 5 involve only finite sets and so hold by construction. Q.E.D.

COROLLARY 3.1: *For $n \geqq 3$ and any finite or countable k, $FL(n)$ has $FL(k)$ as a sublattice.*

THEOREM 7: *$FL(n)$ has no sublattice isomorphic to the free modular or distributive lattice on more than two generators.*

PROOF: Otherwise $FL(n)$ would have $FM(3)$ or $FD(3)$ as a sublattice, for if say it had say $FM(4)$ as a sublattice, then the sublattice generated by the first three of these four generators would satisfy the definition of free modular lattice (this argument assumes well-ordering in the infinite case); a similar argument would have sufficed for Lemma 3.1. But it is known ([7] p. 246; [3] pp. 49, 84) that the generators of $FM(3)$ and $FD(3)$ satisfy the conditions of Theorem 5, so the sublattice they generate would be $FL(3)$ rather than $FM(3)$ or $FD(3)$.

4. Intrinsic topology of $FL(n)$

In a lattice each pair of elements, a and b, must have a least upper and greatest lower bound ([3] p. 16, [14] p. 325). By the associative law this must also hold for any finite number of elements, but it may or may not hold for an infinite number (e.g., the set of all integers is a lattice under the customary linear ordering, but there is no greatest; on the other hand any class of subclasses of a fixed class does have a least upper bound under set inclusion). In case the least upper bound or greatest lower bound of a set $\{a_i\}$ exists we may denote it sup $\{a_i\}$ or inf $\{a_i\}$ respectively ([3] p. 16).

DEFINITION 4.1 ([3] p. 27): In a lattice, $\{a_i\}$ is said to *(o)-converge* to a if sequences $\{u_i\}$ and $\{v_i\}$ exist such that

$$u_i \leqq u_{i+1} \leqq a_{i+1} \leqq v_{i+1} \leqq v_i$$

for all i, and sup $\{u_i\}$ = inf $\{v_i\}$ = a. *Then we write $a_i \to a$, $u_i \uparrow a$, and $v_i \downarrow a$.*

In terms of this order convergence, one may study the topology of a lattice—

the intrinsic topology, as Birkhoff calls it, since it is defined in terms of the inclusion relation without external considerations. For instance, the lattice operations are said to be *continuous*[6] if $a_i \to a$ and $b_i \to b$ imply $a_i \cup b_i \to a \cup b$ and dually. For this, it is sufficient ([3] p. 30) that $a_i \downarrow a$ imply $a_i \cup b \downarrow a \cup b$ and dually. If sup $\{a_i\}$ and inf $\{a_i\}$ exist for every set $\{a_i\}$ then the lattice is said to be *complete*.

We have already (§4 of [14]) obtained a simple topological result for $FL(n)$, in finding that in certain cases one element covers another; thus $FL(n)$ is in a sense not everywhere dense. Moreover, in Theorem 9 below we find that $FL(n)$ is not complete—we exhibit sets without suprema. This leaves open the question of the existence of an infinite ascending chain with a supremum; if there is none then Theorem 8 is vacuous.

THEOREM 8: *$FL(n)$ is continuous.*

PROOF: Suppose $a_i \downarrow a$. By [3] (p. 30) it suffices to show $a_i \cup b \downarrow a \cup b$. But $a \cup b$ is obviously a lower bound to the $a_i \cup b$; it remains to be shown greatest—that is, that if c is any lower bound to the $a_i \cup b$ then $c \leqq a \cup b$. But if this should fail—if $c \not\leqq a \cup b$—then $c \cup a \cup b > a \cup b$ and $c \cup a \cup b$ is again a lower bound, so it suffices to show by induction on $L(c)$ that it is impossible to have $c > a \cup b$ and c a lower bound to the $a_i \cup b$.

CASE 1: $c \equiv \sum_{j=1}^{m} c_j$, where we may suppose that no c_j is a join. Then $c \leqq a_i \cup b$ for all i, so $c_j \leqq a_i \cup b$ for all i and j. Hence by Theorem 1, given i and j, either (1) $c_j \leqq a_i$, or (2) $c_j \leqq b$, or (3) $c_j \equiv \prod_k c_{jk}$ and for some k, $c_{jk} \leqq a_i \cup b$, the possibility (3) being excluded if c_j is a generator. If (3) holds for some j and infinitely many i, it holds for that j and infinitely many i and some fixed k, since m is finite. By monotonicity of the a_i, $c_{jk} \cup \sum_{t \neq j} c_t$ is a shorter lower bound than c and contains $a \cup b$, contrary to hypothesis of induction. Thus, given j, it must be that (1) holds for infinitely many i or (2) does; perhaps both. In the former case $c_j \leqq a$, in the latter $c_j \leqq b$, so $c \equiv \sum c_j \leqq a \cup b$ contrary to hypothesis that $c > a \cup b$.

CASE 2: $c \equiv \prod_{j=1}^{m} c_j$. Then $\prod c_j \leqq a_i \cup b$ for all i, so given i either (1) $c \leqq a_i$, or (2) $c \leqq b$, or (3) for some j, $c_j \leqq a_i \cup b$. If (3) holds for infinitely many i, then (since m is finite) it holds for some fixed j and infinitely many i, and hence for that j and all i, and c_j is a shorter lower bound, $c_j > a \cup b$, contrary to induction. Otherwise (1) or (2) holds for infinitely many i, so $c \leqq a$ or $c \leqq b$, contrary to $c > a \cup b$.

CASE 3: c is a generator. This is included in Case 2, with (3) impossible. This starts the induction. Thus the theorem holds.

This leaves open the question of whether there *are* any infinite sequences which have limits. We proceed to show that at any rate in $FL(3)$ *not every* infinite sequence has a limit. For our horrible example we consider the elements

[6] Not to be confused with the use by some authors of "continuous" for what we call "complete," nor with the use of "continuous" in connection with a metric in a lattice ([3], p. 43).

which Birkhoff used[7] to show that $FL(3)$ has an infinite number of distinct elements.

Let $t_1 \equiv x_1$, $t_{6i+2k} \equiv t_{6i+2k-1} \cap x_{2k}$, $t_{6i+2k+1} \equiv t_{6i+2k} \cup x_{2k+1}$ for $i = 0, 1, 2, \cdots$, where the subscripts on x are taken modulo 3. Thus for instance

$$t_{6i+7} \equiv (((((t_{6i+1} \cap x_2) \cup x_3) \cap x_1) \cup x_2) \cap x_3) \cup x_1 .$$

By Theorem 1, $t_1 \leqq t_7$, and by induction $t_{6h+k} \leqq t_{6i+k}$ in $FL(3)$, if $h < i$. Birkhoff's example just cited shows that equality cannot hold, as could also be shown by Theorem 1 and induction, so $t_{6h+k} < t_{6i+k}$ in $FL(3)$ if $h < i$. We have thus six infinite strictly monotone increasing chains: t_1, t_7, t_{13}, $\cdots$; t_2, t_8, t_{14}, $\cdots$, etc.

LEMMA 4.1: *If one of the infinite chains t_{6i+k} (k fixed) has a least upper bound, then so does each, and if* $\sup_i \{t_{6i+k}\} = s_k$, *then* $s_{2k} = s_{2k-1} \cap x_{2k}$ *and* $s_{2k+1} = s_{2k} \cup x_{2k+1}$.

NOTE: The subscripts on s may be taken modulo 6, on x modulo 3. We do not as yet assert that $s_{2k-1} \cap x_{2k}$ gives the canonical form of s_{2k}.

PROOF: By Theorem 8, if s_{2k-1} exists then

$$\begin{aligned} s_{2k} \equiv \sup_i \{t_{6i+2k}\} &\equiv \sup_i \{t_{6i+2k-1} \cap x_{2k}\} \\ &= \sup_i \{t_{6i+2k-1}\} \cap \sup_i x_{2k} \\ &\equiv s_{2k-1} \cap x_{2k} \end{aligned}$$

so that s_{2k} exists and equals $s_{2k-1} \cap x_{2k}$. Likewise $s_{2k+1} = s_{2k} \cup x_{2k+1}$ and so if any s exists, then by repetition all do. Q.E.D.

LEMMA 4.2: *If the s_k are in canonical form, then the equalities in Lemma* 4.1 *become identities.*

PROOF: PART (I): $s_{2k+1} = s_{2k} \cup x_{2k+1}$.

CASE 1: $s_{2k+1} \equiv \prod r_i$. Then either $s_{2k+1} = s_{2k}$ or $s_{2k+1} = x_{2k+1}$, by Corollary 1.2. But the second is contradicted by $x_{2k+1} \leqq t_{6i+2k+1} < s_{2k+1}$, and if the former held, then by Lemma 4.1,

$$x_{2k+1} < s_{2k+1} = s_{2k} \leqq x_{2k}$$

whereas $x_{2k+1} \not\leqq x_{2k}$ in $FL(3)$. Thus Case 1 is impossible.

CASE 2: $s_{2k+1} \equiv \sum r_j$. By Corollary (19) of [14] and by (I), we have

(1) given j, either $r_j \leqq s_{2k}$ or $r_j \leqq x_{2k+1}$.

Now $x_{2k+1} \leqq s_{2k+1}$, so by Theorem 1 B, $x_{2k+1} \leqq r_j$ for some j; say

(2) $$x_{2k+1} \leqq r_1 .$$

But as in Case 1,

(3) $$x_{2k+1} \not\leqq s_{2k} .$$

[7] [2] pp. 451–452. The following typographical changes are needed on p. 452: line 5, subscript should be $i + 1$; lines 7–8, interchange meet and join symbols.

By (2) and (3), $r_1 \not\leqq s_{2k}$, so by (1) and (2), $r_1 = x_{2k+1}$. By canonicity and (1), we have

(4) $\qquad r_1 \equiv x_{2k+1} \quad$ and $\quad r_j \leqq s_{2k} \quad$ for $\quad j > 1.$

By Lemma 4.1,

$$s_{2k-1} \cap x_{2k} = s_{2k} \leqq s_{2k+1} \equiv \sum r_j,$$

so by Theorem 1, either

(5) $\qquad s_{2k} \leqq r_i \quad$ for some i, $\quad$ or

(6) $\qquad s_{2k-1} \leqq s_{2k+1} \quad$ or

(7) $\qquad x_{2k} \leqq s_{2k+1}.$

If (7) holds, then $x_{2k} \cup x_{2k+1} \leqq s_{2k+1}$, which is impossible in $FL(3)$ by the covering results stated in Theorem 3 since s_{2k+1} is the limit of a strictly monotone increasing sequence. If (6) holds then a similar contradiction is reached: $x_{2k-1} \cup x_{2k+1} \leqq s_{2k+1}$. Hence (5) holds. But $s_{2k} \not\leqq x_{2k+1}$, for

$$x_{2k-1} \cap x_{2k} \leqq (s_{2k-2} \cup x_{2k-1}) \cap x_{2k} = s_{2k-1} \cap x_{2k} = s_{2k}$$

whereas $x_{2k-1} \cap x_{2k} \not\leqq x_{2k+1}$. Hence in (5) $i > 1$ and by (4), $r_i = s_{2k}$. By canonicity,

$$s_{2k+1} \equiv r_1 \cup r_2 \equiv x_{2k+1} \cup s_{2k}.$$

Case 3: $s_{2k} \equiv x_i$. Like Case 1.

Part (II): $s_{2k} = s_{2k-1} \cap x_{2k}$.

Case 1: $s_{2k} \equiv \sum r_j$. By Corollary 1.2, either

(1) $\qquad s_{2k} = s_{2k-1} \quad$ or

(2) $\qquad s_{2k} = x_{2k}.$

If (1) holds, then

$$x_{2k-1} \leqq s_{2k-1} = s_{2k} \leqq x_{2k}$$

whereas $x_{2k-1} \not\leqq x_{2k}$. If (2) holds, then

$$x_{2k} \cup x_{2k+1} \leqq s_{2k} \cup x_{2k+1} = s_{2k+1}$$

contrary to Theorem 3.

Case 2: $s_{2k} \equiv \prod r_j$. By Corollary (19) of [14], given j, either

(1) $\qquad s_{2k-1} \leqq r_j \quad$ or

(2) $\qquad x_{2k} \leqq r_j.$

Also $s_{2k} \leqq x_{2k}$, so for some j, $r_j \leqq x_{2k}$; say $j = 1$:

(3) $\qquad r_1 \leqq x_{2k}.$

Then $s_{2k-1} \not\leqq r_1$ (otherwise $x_{2k-1} \leqq s_{2k-1} \leqq r_1 \leqq x_{2k}$, whereas $x_{2k-1} \not\leqq x_{2k}$) so by (2), $x_{2k} \leqq r_1$, and by (3) and canonicity we have

$$x_{2k} \equiv r_1 \quad \text{and} \quad s_{2k-1} \leqq r_j \quad \text{for} \quad j > 1. \tag{4}$$

Also

$$\prod r_i \equiv s_{2k} = s_{2k-1} \cap x_{2k} \leqq s_{2k-1} = s_{2k-2} \cup x_{2k-1},$$

so either

$$s_{2k} \leqq s_{2k-2} \quad \text{or} \tag{5}$$

$$s_{2k} \leqq x_{2k-1} \quad \text{or} \tag{6}$$

$$r_i \leqq s_{2k-1} \quad \text{for some } i. \tag{7}$$

If (5) holds, then $s_{2k} \leqq x_{2k-2} \cap x_{2k}$, and if (6) holds then $s_{2k} \leqq x_{2k-1} \cap x_{2k}$, each contrary to Theorem 3. Hence (7) holds, and by (4) and canonicity,

$$s_{2k} \equiv x_{2k} \cap s_{2k-1}.$$

CASE 3: $s_{2k} \equiv x_i$. Like Case 1.

Thus Lemma 4.2 is proven.

LEMMA 4.3: *$FL(3)$ is not complete.*

PROOF: The assumption that every set has a least upper bound leads via Lemmas 4.1 and 4.2 (applied six times) to the conclusion that

$$s_1 = (((((s_1 \cap x_2) \cup x_3) \cap x_1) \cup x_2) \cap x_3) \cup x_1$$

and that both sides of this equation are in canonical form, which is impossible since they are not of the same length.

THEOREM 9: *$FL(n)$ is not complete*[8], for $n \geqq 3$.

PROOF: $FL(n)$ has the infinite chain $\{t_{6i+1}\}$, of which each element is a polynomial in x_1, x_2, x_3 alone, not involving $x_4, \cdots, x_n$. By[9] Lemma 3.2, the least upper bound of this chain (if it exists) is a polynomial in x_1, x_2, x_3 alone, and so by Lemma 3.1 must be the least upper bound in $FL(3)$, contrary to the proof of Lemma 4.3.

5. $FM(4)$

Birkhoff has shown ([1] p. 463–464) that the free modular lattice $FM(4)$ has infinitely many distinct elements. A slightly stronger result, again due to Birkhoff[10] is this:

THEOREM 10: *$FM(4)$ has an infinite chain of distinct elements.*

PROOF: Let A be the free Abelian group generated by $x_1, x_2, \cdots, x_n, \cdots$. For $k = 0, 1, 2, \cdots$ let $S_1 = [x_{2k}]$, the subgroup of A generated by the x_{2k}, $S_2 = [x_{2k} + x_{2k+1}]$, $S_3 = [x_{2k+1}]$, $S_4 = [x_{2k+1} + x_{2k+2}]$. If we set $B = [S_1, S_2,$

[8] In fact, not even conditionally sigma-complete ([3] p. 29); we have exhibited a bounded countable set without least upper bound. Note: While n was required finite in Theorem 3, the method of applying that theorem here avoids such a restriction in the present theorem.

[9] The adjunction of 1 and 0 possibly required for that lemma causes no trouble, for (using Lemma 3.1) we may apply Lemma 3.2 to the lattice generated by the first three x's and those involved in an alleged least upper bound; since this set is finite, 1 already exists for this sublattice.

[10] Who communicated this section to the author by letter, March 7, 1940.

S_3, $S_4]$, then the set of all subgroups of B is a modular lattice ([3] p. 35). Let $T_1 = S_1$, $T_{4n+1} = T_{4n} \cap S_1$, $T_{4n+2} = T_{4n+1} \cup S_2$, $T_{4n+3} = T_{4n+2} \cap S_3$, $T_{4n+4} = T_{4n+3} \cup S_4$. Then

$$T_{4n} = [x_{2n+2k-1}, x_{2n+2k}, x_{2k+1} + x_{2k+2}]$$
$$T_{4n+1} = [x_{2n+2k}]$$
$$T_{4n+2} = [x_{2n+2k}, x_{2n+2k+1}, x_{2k} + x_{2k+1}]$$
$$T_{4n+3} = [x_{2n+2k+1}],$$

whence $T_4 > T_8 > T_{12} > \cdots$. Thus one modular lattice with four generators has an infinite chain, hence so does $FM(4)$. Q.E.D.

6. Unsolved problems

As far as the author is aware, the following problems suggested by or related to the present paper remain open. (1) Solve the word problem for free modular lattices ([3] p. 146). This seems to be difficult. (2) Extend Theorem 3, and [14] §4. Perhaps not difficult. (3) Does *some* infinite set in $FL(n)$ order-converge (§4)? We may call a set of elements $\{a_1, a_2, \cdots\}$ a *fence* below a if $c < a$ implies $c \leqq a_i$ for some i. As we saw in Theorem 3 the $\sum_{i \neq p} x_i$ form a fence below the unit in $FL(n)$; etc. Is it perhaps true that in $FL(n)$ *every* element has a finite upper and lower fence? This would settle the first part of (3). On the other hand if some suprema exist, are there real cases like the vacuous one of Lemma 4.2 where canonical forms carry over? Do there exist infinite connected chains in $FL(n)$—ones in which no further elements can be interpolated? (4) Study the free complete lattice. Study the completion of $FL(n)$ by cuts (MacNeille [9] p. 443; [3] p. 27). (5) Determine the finite sublattices of $FL(n)$; cf. Theorem 7.

HARVARD UNIVERSITY AND THE UNIVERSITY OF PENNSYLVANIA

REFERENCES

G. BIRKHOFF. [1]: *On the combination of subalgebras*, Proc. Camb. Phil. Soc. 29 (1933) pp. 441–464. [2]: *On the structure of abstract algebras*, ibid., 31 (1935) pp. 433–454. [3]: *Lattice Theory*, Amer. Math. Soc. Colloq. Pub. XXV, New York 1940. [4]: *Rings of Sets*, Duke J. 3 (1937) pp. 443–454.

R. CHURCH. [5]: *Numerical analysis of certain free distributive structures*, Duke J. 6 (1940) pp. 732–734.

R. DEDEKIND. [6]: *Über Zerlegungen von Zahlen durch ihre grossten gemeinsam Teiler*, Festschrift Techn. Hoch. Braunschweig (1897), Gesammelte Werke II pp. 103–148. [7]: *Über die von drei Moduln erzeugte Dualgruppe*, Math. Annalen 53 (1900) pp. 371–403, Ges. Werke II pp. 236–271.

F. LEVI. [8]: *Über die Untergruppen freier Gruppen*, Math. Zeit. 32 (1930) p. 315ff.

H. M. MACNEILLE. [9]: *Partially ordered sets*, Trans. Am. Math. Soc. 42 (1937) pp. 416–460.

J. NIELSEN. [10]: *Om regning med ikke-kommutative faktorer*, Matematisk Tidsskrift B, 1921, pp. 77–94.

O. SCHREIER. [11]: *Die Untergruppen der freien Gruppen*, Abh. aus den Math. Seminar Hamburg 5 (1927) pp. 161–183.

T. SKOLEM. [12]: *Über gewisse Verbände oder Lattices*, Avhandlinger utgitt av det Norske Videnskaps Akademi (Mat.-Naturv. Klasse) 1936 no. 7.

P. M. WHITMAN. [13]: Abstract, Bull. Am. Math. Soc. 46 (1940) p. 760. [14]: *Free Lattices*. Annals of Math. (2) 42 (1941) pp. 325–330.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

TOPOLOGICAL METHODS FOR THE CONSTRUCTION OF TENSOR FUNCTIONS

BY NORMAN E. STEENROD

(Received July 7, 1941)

1. Introduction

It is a well-known theorem that an orientable manifold M admits a continuous field of non-zero tangent vectors if and only if the Euler number of M is zero. Stiefel [5][1] has generalized this result to the case of any finite number of fields independent at each point, obtaining necessary conditions in the form of the vanishing of certain cohomology classes of M.

From these results, there is every reason to expect a general theory connecting the existence of tensor functions of various types over a manifold M with the topological structure of M. It is our purpose in this paper to give the foundations and first theorems of such a theory.

Briefly, it is shown that the set M' of point tensors (of a fixed order and weight) over the differentiable manifold M is a differentiable manifold forming, in a natural way, a fibre bundle over M in the sense of Whitney [6]. A tensor function attaches to each point of M a point of the fibre in M' over it. The nature of a tensor function is restricted by specifying that its values lie in a submanifold M'' of M'. If M'' is likewise a fibre bundle, then there is a characteristic cohomology class in M attached to M''. The vanishing of this class proves to be a necessary (and sometimes sufficient) condition for the existence of the prescribed tensor function. As an application, a direct and simple proof is given that any separable manifold admits a Riemann metric.[2]

It is worth noting here that the characteristic cohomology class belongs to a new type of cohomology group: one based on local coefficient groups connected by local isomorphisms (see section 10). It reduces to the usual group if M is simply connected.

It will be seen in the proofs that the problem we are considering is a generalization of the problem of extending continuous mappings. We shall both generalize and use extensively certain theorems of Eilenberg in this connection [2].

In an appendix, the existence and properties of the characteristic cohomology class are established for the more general type of fibre space introduced by Hurewicz and the author [3]. Any lengthy proof that would interrupt the trend of ideas is postponed to a later section.

---

[1] Numbers in square brackets refer to the bibliography.

[2] Whitney [7] has proved this by showing that any C^r-manifold is C^r-homeomorphic to an analytic manifold in a euclidean space.

2. Spaces of point tensors

In the following, M will denote the differentiable manifold[3] of dimension n over which tensors are to be defined. The class r of M is assumed to be $\geqq 2$. It is not assumed that M is compact. It is assumed that M is separable so that a countable set of coördinate neighborhoods can be found to cover it. It follows from results of Cairns [1] that M can be triangulated. We may therefore assume that M is homeomorphic with a definite simplicial countable, complex. The symbol M^q will denote the subcomplex of M composed of its simplexes of dimensions $\leqq q$.

In order to be explicit and yet not have too cumbersome a notation, we shall define only the manifold M' of point tensors over M having one contravariant and two covariant indices. If P is a point of M, $C(x)$ a coördinate neighborhood of P, and a^i_{jk} $(i, j, k = 1, \cdots, n)$ an ordered set of n^3 real numbers, the combination $(P, C(x), a^i_{jk})$ is called a *representation of a point tensor*. Two such $(P, C(x), a^i_{jk})$, $(Q, C(\bar{x}), \bar{a}^i_{jk})$ are said to be equivalent if $P = Q$ and

$$\text{(A)} \qquad \bar{a}^i_{jk} = a^u_{vw} \frac{\partial x^v}{\partial \bar{x}^j} \frac{\partial x^w}{\partial \bar{x}^k} \frac{\partial \bar{x}^i}{\partial x^u}.$$

(The derivatives are evaluated at P.) As is well known, this is a proper equivalence relation, so that the combinations $(P, C(x), a^i_{jk})$ fall into mutually exclusive equivalence classes. An equivalence class is called a *point tensor*. The adjective *point* is used for the obvious reason that a point tensor is attached to that point of M occurring in any one of its representations.

The family of all point tensors of the above type at all points of M form a set M' which we now proceed to show is a differentiable manifold in a natural way. To do this we must define 1–1 maps of subsets of M' into suitable subsets of a euclidean space. Let $C(x)$ be an admissible coördinate neighborhood in M. Let U be the family of all point tensors having representations $(P, C(x), a^i_{jk})$ (where P varies in $C(x)$, and the a's are arbitrary). Then U is in $1-1$ correspondence with this set of representations in $C(x)$. Attach to $(P, C(x), a^i_{jk})$ the $n + n^3$ real numbers $x^1, \cdots, x^n, a^i_{jk}$ where the x's are the coördinates of P in $C(x)$. The two correspondences define a $1-1$ map of U into the open subset of $(n + n^3)$-euclidean space of coördinates (x, a) where x is in $C(x)$ and the a's are arbitrary. Thus for each $C(x)$ in M a coördinate neighborhood $C(x, a)$ in M' is determined. If a point tensor is in both $C(x, a)$ and $C(\bar{x}, \bar{a})$ then, if $\bar{x}^i = f^i(x)$ is the coördinate transformation from $C(x)$ to $C(\bar{x})$, these equations together with the equations (A) define the coördinate transformation from $C(x, a)$ to $C(\bar{x}, \bar{a})$. Thus, the $\bar{x}$'s are functions of the x's alone while the $\bar{a}$'s are functions of both the x's and the a's and are linear in the latter. The determinant of this transformation is easily seen to be a power of the determinant of the transformation from $C(x)$ to $C(\bar{x})$. Since the derivatives in (A) are of class

[3] For the concept of differentiable manifold see [7].

$r - 1$, it follows that M' is a differentiable manifold of class $r - 1$. Its dimension is $n + n^3$.

It is clear that the relative point tensors of a given weight likewise form a differentiable manifold. The same procedure likewise carries through for the space of point affine connections. It is only necessary that the transformation law in question shall define a proper equivalence relation among the representations. For example, ordered pairs of covariant vectors at points of M have representations of the form $(P, C(x), a_i, b_i)$ with coördinate transformations of the form

$$\bar{x}^i = f^i(x), \qquad \bar{a}_i = a_u \frac{\partial x^u}{\partial \bar{x}^i}, \qquad \bar{b}_i = b_u \frac{\partial x^u}{\partial \bar{x}^i}.$$

The determinant is again a power of the determinant of $x \to \bar{x}$.

In the following pages we shall frequently speak of "a tensor manifold M' over M". It is to be understood that M' is the space of point tensors over M of a fixed order and weight—or M' is the space of point affine connections over M—or M' is the space of ordered sets of vectors (or tensors of given order and weight) at points of M.

3. The projection of the tensor manifold onto M

Let the function π assign to each point tensor in M' the point of M to which the tensor is attached. This natural mapping of M' onto M we refer to as the *projection*. If $C(x)$ is a coördinate neighborhood in M, and $C(x, a)$ the corresponding one in M', then π has the form of a projection in these coördinates. The Jacobian of the projection has rank n at each point of M.

The inverse image of a point P of M in M' (i.e. the point tensors at P) form a linear space which we shall refer to as the fibre F'_P over P. It follows readily that M' is a fibre bundle over M in the sense of Whitney [6]. The fibres are euclidean spaces with linear transformations as the admissible group, and the fibres over a particular coördinate neighborhood form a product space.

4. Tensor functions over M

A *tensor function* over M is a map f of M in a tensor manifold M' over M having the property that πf is the identity. The image of M in M' under f is called the graph of f.

The usual definition of tensor function has the disadvantage that graphs exist only locally and then vary with the coördinate system. The advantage of the present method is that, by identifying the equivalent functional values first, the function itself falls into that class of objects (usually referred to as functions) which are point to point correspondences.

If f is a map of the set A in the set B, the graph of f is usually defined to be the set $(a, f(a))$ in the product space $A \times B$. Therefore the foregoing definition of graph of a tensor function needs a few words of justification. Suppose M' is the space of point scalars over M. It is not hard to see that M' is the product

space $M \times L$ of M with a coördinate line L. Any map f of M in L (i.e. a scalar function in the usual sense) has a graph in $M \times L$ which defines a scalar function in the sense described above. Therefore, at least in this case, the term is justified. An additional reason is that for any neighborhood $C(x)$ the part of the graph over it is the graph of the functions $a(x)$ in the product space $C(x, a)$.

In general, as examples given later will show, M' is not the product space of M with a fibre. Consequently we cannot expect that tensor functions as usually defined will have graphs in the usual sense. We have, however, with the present definition, most of the important features of a graph. The graph is in a space which is locally a product space over M, and we have the projection π which, when applied to the graph, results in the identity map of M. The only missing feature is the resolution of M' into a product space of a second space with M. This, however, is inherent in the nature of tensor.

In case M can be resolved into a product space, it is of little interest unless the resolution satisfies several strong conditions. We shall say that M' is *properly resolved* into a product space $M \times E$ where E is a euclidean space provided there is a homeomorphic map of class $r - 1$ of $M \times E$ onto M' with the property that, for each $P \in M$, the section $P \times E$ is mapped linearly on the fibre over P in M. Then we have

THEOREM 1. *If M' is a tensor space over M of dimension $n + m$, it can be properly resolved into a product space if and only if there exist m tensor functions over M of class $r - 1$ which are linearly independent at each point of M.*

Suppose M is properly resolved, and ϕ is its homeomorphism with $M \times E$. Let $a_1, \cdots, a_m$ be m linearly independent points of E. Then $f_i(P) = \phi(P, a_i)$ $(i = 1, \cdots, m)$ are m tensor functions of class $r - 1$ which are linearly independent at each point of M.

Conversely, let $\{f_i(P)\}$ be m independent tensor functions of class $r - 1$. Define $\phi(P, a_1, \cdots, a_m) = \sum a_i f_i(P)$. Then ϕ is a proper map of $M \times E$ onto M.

5. Statement of problem

The most general type of problem for which we shall give a method of attack may be formulated as follows:

Let M' be a tensor manifold over M, and let M'' be a differentiable submanifold of M' such that the projection π maps M'' onto M, has a Jacobian of rank n at every point of M'', and M'' is locally a product space over M. (This last means that each point P of M has a neighborhood C such that the product $C \times F''$, where F'' is the fibre over P, is homeomorphic with the part of M'' over C, and, for any point Q of C, $Q \times C$ is mapped on the fibre over Q.) The problem is to define a map f of M in M'' such that πf is the identity map of M, and f has a specified class (up to the class of M'').

The following are examples of problems of this type.

a. If M'' is all of M', the solution is immediate, for the zero tensor at every point is an f of the required kind.

b. Let M' be the space of contravariant point vectors over M and M'' the family of non-zero vectors. This is the problem of defining a non-singular vector field over M.

c. Let M' be the covariant point tensors of order 2, and M'' the symmetric positive definite ones. A solution f determines a Riemann metric in M.

d. Let M' be as in c, and let M'' be the symmetric semi-definite tensors of a fixed rank k.

e. Let M' be as in c, and let M'' be the symmetric non-singular tensors of a fixed index h.

f. Let M' be the set of ordered sets of k contravariant vectors, $(k \leqq n)$, and let M'' be the subspace of independent sets. This is the problem considered by Stiefel [5].

g. Consider the problem raised by Theorem 1. Let M'' be the tensor space of ordered sets of m tensors each of the same order as in M'. Let M''' be the subspace of independent sets. Then the existence of a map f of M into M''' such that πf = identity is a necessary and sufficient condition that M' can be properly resolved into a product space.

The methods to be given seem to be of little value in handling a problem involving differential conditions such as defining a flat affine connection or a Riemann metric of constant or constant mean curvature.

The following theorem reduces the general problem to the simpler one of finding a continuous admissible tensor defined on M.

THEOREM 2. *If f is a continuous map of M in M'' such that πf = identity, then there exists a map f' of class = that of M'' such that $\pi f'$ = identity. If a metric is given in M'', f' may be chosen to approximate f as closely as desired. (If M'' is analytic, we can only assert the existence of an f' of class ∞).*

6. Construction of a Riemann metric

In order to illustrate the basic method of attack on our problem, we shall show that one can define over M a symmetric, positive definite, tensor function g of covariant order 2. This result has already been established by Whitney [7]. His procedure is to construct a regular imbedding of M in a euclidean space and then to take over into M the metric of the euclidean space. Our procedure is a direct construction of the metric tensor and is somewhat simpler. However, it fails to give the imbedding theorem.

Let M' be the manifold of covariant point tensors of order 2 over M. Let M'' be the subset of symmetric, positive definite ones. Let F' be the fibre in M' over a point P in M, and let F'' be its part in M''. The important observation at this point is that F'' is a convex linear cell of dimension $n(n+1)/2$. This is proved by noting that a linear combination with positive coefficients of two positive definite quadratic forms is positive definite. Since a symmetric matrix is determined by $n(n+1)/2$ of its elements, and since any symmetric matrix sufficiently near a positive definite one is likewise positive definite, it follows that the dimension is $n(n+1)/2$.

We suppose now that M is subdivided into a simplicial complex so fine that each simplex is contained wholly within some one coördinate neighborhood. Since the number of simplexes is countable, we may order them in a sequence $\sigma_1, \sigma_2, \cdots$ in such a way that a simplex is preceded by any one of its faces. Then σ_1 is a vertex P of M. We choose a point in the fibre F'' over P and denote it by $g(P)$. Suppose inductively that g has been defined over all simplexes preceding σ_k so as to be continuous on their point set sum and πg = identity. Then g is defined on $\partial\sigma_k$ (the boundary of σ_k). Let $C(x)$ be a coördinate neighborhood containing σ_k. Then the functions $g_{ij}(x)$ are defined and continuous over $\partial\sigma_k$. These functions alone map $\partial\sigma_k$ into the fibre F'' over a point P in $C(x)$. Since F'' is a cell, these functions can be extended over the interior of σ_k so as to give a continuous map of the closed simplex σ_k in F''. The desired map g of σ_k in $C(x, g_{ij})$ is now obtained by setting $g(x) = (x, g_{ij}(x))$ for $x \in \sigma_k$. Then g is continuous on the sum of the first k simplexes, for if a function is continuous on each of two compact sets, it is continuous on their sum. This inductive construction leads to a continuous map g of M in M''. The differentiable approximation is given by Theorem 2.

7. Observations on the construction

It is clear that the construction just given applies in any problem where F'' is a cell. This condition may be weakened to the extent of requiring that every continuous map of a $(k-1)$-sphere in F'' for $k \leqq n$ is contractible to a point in F''. This latter condition is equivalent to the requirement that all homotopy groups of F'' of dimensions $<n$ shall vanish. This in turn is equivalent to the same requirement on the homology groups of F'' of dimensions $<n$ [4].

As an application, suppose it is required to define a tensor function over M of a prescribed order >1 which is not zero at any point. In this case the fibre F'' is a euclidean space with its origin deleted. Since the order is >1, the dimension m of the euclidean space is $>n$. Deleting a single point from a euclidean m-space does not alter the fact that a map of a sphere of dimension $<m-1$ in the space is contractible to a point. The required function therefore exists.

We formulate our results as follows:

THEOREM 3. *If all the homotopy groups or equally well all the homology groups of the fibre F'' vanish for each dimension $<n$, then there exists a tensor defined over M of the specified type.*

8. Construction of the characteristic cocycle

If some sphere in F'' of dimension $<n$ is not contractible, we can expect essential difficulties. In order to discuss this case we make the

ASSUMPTIONS. Let h be the smallest integer such that the homotopy group [4], $\pi_h(F'') \neq 0$. We suppose $h < n$. By $\pi_0(F'')$ is meant the 0^{th} homology group of F'' where cycles are taken with integer coefficients, and the sum of the coefficients of a cycle is zero. If $h = 1$, we shall suppose that $\pi_1(F'')$ is abelian.

Since M'' is locally a product space, any two nearby fibres are homeomorphic. As M is connected we can pass from one fibre to any other through a succession of fibres such that successive pairs are homeomorphic. Thus, the above assumptions need only have been made of one fibre.

It is a consequence of our assumptions that any map of an oriented h-sphere in F'' determines a unique element of $\pi_h(F'')$ (i.e. it is not necessary that a fixed reference point of the sphere be mapped on a fixed reference point of F'').

We shall suppose that M is subdivided so fine that the star of a simplex lies wholly within a single admissible neighborhood of M (one such that the part of M'' over it is a product $C\times F''$). Then, by the argument of section 7, one may construct a map f of the h-dimensional skeleton M^h in M'' such that $\pi f =$ identity.

To any map f of M^h in M'' such that $\pi f =$ identity, we attach a chain $c^{h+1}(f)$ of M as follows. If σ is an oriented $(h+1)$-simplex, C an admissible neighborhood containing σ, and $C\times F''$ a representation of $\pi^{-1}(C)$ in M'', then f maps $\partial\sigma$ into $C\times F''$. Let λ map $C\times F''$ into F'' by attaching to each point its F'' coördinate. Then λf maps $\partial\sigma$ into F'', determining thereby an element of $\pi_h(F'')$ which we denote by $c(f, \sigma)$. Then[4]

$$c^{h+1}(f) = \sum c(f, \sigma^{h+1})\sigma^{h+1}.$$

The sum extends over all $(h+1)$-simplexes of M.

An object is a *chain* if it is a function from simplexes to a group. Here F'' varies from one neighborhood to another, so the definition of chain is not satisfied by $c^{h+1}(f)$. The definition will be satisfied when we have established a fixed reference F_0'', and fixed isomorphisms connecting $\pi_h(F'')$ for each F'' to $\pi_h(F_0'')$. We know such isomorphisms exist; but arbitrarily chosen ones will not serve our purpose. We consider this matter in the following two sections.

9. The orientable case

Let A be a curve joining two points P, Q of M defined by a function $\phi(t)$, $a \leqq t \leqq b$, $\phi(a) = P$, $\phi(b) = Q$. Suppose a continuous function $h(y, t)$ is given with values in M'', defined for y in F_P'' and $a \leqq t \leqq b$, such that $\pi h(y, t) = \phi(t)$ and $h(y, 0) = y$. Then we shall say that h *deforms* F_P'' *along* A *into* F_Q''. Two such deformations of F_P'' are said to be *equivalent* if the two resulting maps of F_P'' in F_Q'' are homotopic in F_Q''.

If the curve A lies in a neighborhood C and the part of M'' over C is represented as a product $C\times F''$, then $h(y, t)$ can be defined as the point $(\phi(t), y)$ of the product $C\times F''$. Now any curve A can be expressed as a sum of curves $A_1, \cdots, A_k$ each of which lies in a single neighborhood. By piecing together homotopies along each, one obtains a homotopy along A.

A homotopy of F_P'' along A into F_Q'' induces an isomorphism between $\pi_h(F_P'')$

[4] The chain $c^{h+1}(f)$ is the "characteristic cocycle" to be found in the work of Stiefel [5] and Whitney [6].

and $\pi_h(F''_Q)$. This isomorphism appears to depend on several factors. However, we have

LEMMA 1. *If A_1, A_2 are two curves from P to Q which are homotopic leaving their end points fixed, then a deformation of F''_P along A_1 into F''_Q is equivalent to one along A_2.*

The virtue of the lemma is that it assures us that the isomorphism set up between $\pi_h(F''_P)$ and $\pi_h(F''_Q)$ by deforming F''_P along a curve into F''_Q depends only on the homotopy class of the curve. Consider now the case $P = Q$. Then every element of $\pi_1(M)$ defines an automorphism of $\pi_h(F''_P)$. It is not hard to see that the attached automorphism gives a homomorphism of $\pi_1(M)$ into the group of automorphisms of $\pi_h(F''_P)$.

DEFINITION. We say that M'' is *orientable* over M relative to $\pi_h(F'')$ if, for each point P of M and each element of $\pi_1(M)$, the resulting automorphism of $\pi_h(F''_P)$ is the identity.

The use of the word *orientable* is explained by the fact that M itself is orientable if and only if the space M'' of non-zero, contravariant, point vectors is orientable over M relative to $\pi_{n-1}(F'')$.

The next remark is that in the definition of orientability we do not need to require of each point P that any element of $\pi_1(M)$ shall induce the identity automorphism. It is sufficient to demand this of just one point P_0; it then follows for every point P. This follows from the well-known fact that any closed curve beginning and ending at P is homotopic to a path which first describes a curve A from P to P_0, then a closed curve from P_0 to P_0, then finally describes A^{-1}.

The virtue of the orientable case is that any two deformations of F''_P along curves into F''_Q induce the same isomorphism between $\pi_h(F''_P)$ and $\pi_h(F''_Q)$. For, if A, B are two curves from P to Q, and ϕ_1, ϕ_2 are the two corresponding isomorphisms, then deforming F''_Q around the closed curve $A^{-1}B$ gives the identity isomorphism of $\pi_h(F''_Q)$. This automorphism applied to ϕ_1 gives ϕ_2, since the path $AA^{-1}B$ is homotopic to B.

Assuming now the orientable case, we choose a fixed fibre F''_0, and for any fibre F'' we set up the unique isomorphism between $\pi_h(F'')$ and $\pi_h(F''_0)$ obtained by deforming F'' along a curve into F''_0. In this way, each element of $\pi_h(F'')$ for any F'' represents a unique element of $\pi_h(F''_0)$. Consequently, in the orientable case, the object $c^{h+1}(f)$ defined above is a chain (in the proper sense of the word) with coefficients in the group $\pi_h(F''_0)$.

10. The non-orientable case

Here we cannot proceed as before since no unique isomorphisms in the large exist. However, for each simplex σ, unique isomorphisms connecting the $\pi_h(F''_P)$ for $P \in \sigma$ can be set up using curves in σ; for a simplex is simply connected. Thus each simplex σ of M has a coefficient group $G(\sigma)$. A q-chain is now defined to be a function f attaching to each q-simplex σ an element $f(\sigma) \in G(\sigma)$. The q-chains form a group in the obvious way. Suppose σ' is a face of σ ($\sigma' < \sigma$),

then by using a curve in the closure of σ, joining a point of σ to a point of σ', a unique isomorphism $h_{\sigma'\sigma}$ of $G(\sigma)$ onto $G(\sigma')$ can be defined. If $[\sigma:\sigma']$ is the incidence number, we can now define the boundary of a q-chain f to be the $(q-1)$-chain ∂f having on σ' the value

$$\partial f(\sigma') = \sum_{\sigma} [\sigma:\sigma'] h_{\sigma'\sigma}(f(\sigma))$$

which is an element of $G(\sigma')$. Since the closure of σ is simply connected, $\sigma'' < \sigma' < \sigma$ implies $h_{\sigma''\sigma'}h_{\sigma'\sigma} = h_{\sigma''\sigma}$. From this it follows that $\partial\partial f = 0$. Coboundary is defined by

$$\delta f(\sigma) = \sum_{\sigma'} [\sigma:\sigma'] h^{-1}_{\sigma'\sigma}(f(\sigma'));$$

and, again, $\delta\delta f = 0$. Thus cycles, cocycles, homology and cohomology can be defined as usual.[5] Thus, *in the non-orientable case, we are dealing with the local coefficient groups* $\pi_h(F''_P)$ *connected by the local isomorphisms gotten by deforming the* F''_P *along curves.*[6]

11. Properties of the characteristic cocycle

Having agreed on the sense in which $c^{h+1}(f)$ can be regarded as a chain in M, we are prepared to state its principal properties.

THEOREM 4. *Under the assumptions of section* 8, *there exist maps* f *of the* h*-dimensional skeleton* M^h *of* M *in* M'' *such that* πf = *identity. Any such* f *defines a chain* $c^{h+1}(f)$ *with the following properties*:

(a). $c^{h+1}(f) = 0$ *is a necessary and sufficient condition that* f *can be extended continuously to* M^{h+1}, *preserving* πf = *identity.*

(b). $c^{h+1}(f)$ *is a cocycle.*

(c). *If* f' *is any other map of* M^h *in* M'' *such that* $\pi f'$ = *identity, then* $c^{h+1}(f') \backsim c^{h+1}(f)$.

(d). *If* c^{h+1} *is a cocycle cohomologous to* $c^{h+1}(f)$, *there exists a map* f' *of* M^h *in* M'' *such that* $\pi f'$ = *identity, and* $c^{h+1} = c^{h+1}(f')$.

(e). *A necessary and sufficient condition that there be a map* f *of* M^{h+1} *in* M'' *such that* πf = *identity is that, for any such map* f' *of* M^h, *we have* $c^{h+1}(f') \backsim 0$.

---

[5] The resulting groups may be handled much as the ordinary ones. Care must be taken that simplicial transformations are accompanied by local coefficient group isomorphisms. The usual proof of topological invariance carries through with just these modifications. These new groups may also be defined for a general space. Here each point P has its coefficient group $G(P)$, and a homotopy class of curves from P to Q determine an isomorphism which is transitive under addition. It has come to the author's attention that "local coefficients" and Reidemeister's "Überdeckungen" (*Topologie der Polyeder*, Leipzig, 1938) are equivalent concepts. A discussion of the connection and an analysis of this homology theory will be found in a forthcoming paper.

[6] The procedure of Whitney [6], in the case of non-orientable sphere-bundles, is to reverse the sign of certain incidence numbers in M. His method can apply in our case only when the sole automorphism on $G(\sigma)$ is a reversal of sign.

(f). *The cohomology class of* $c^{h+1}(f)$ *is independent of the subdivision of* M *which is used and of the map* f *of* M^h; *it is therefore a topological invariant of the pair of spaces* M, M'' *and the map* π.

12. Proof[6a] of Theorem 2

In order to construct the differentiable approximation, we shall prove some preliminary results.

(A). Let D, D' be two rectangular domains in n-space E^n defined by $a_i < x_i < b_i$, $a_i' < x_i < b_i'$, respectively, and such that D' contains the closure $\bar{D}$ of D. Then there exists a real-valued function g defined in E^n of class ∞, and such that

$$0 \leqq g \leqq 1, \qquad g = \begin{cases} 1 & \text{in } D, \\ 0 & \text{in } E^n - D'. \end{cases}$$

Let (c, d) be an interval and let

$$\psi_{cd}(x) = \begin{cases} \exp\left(-\dfrac{1}{x-c} + \dfrac{1}{x-d}\right) & \text{in } (c, d), \\ 0 & \text{outside } (c, d). \end{cases}$$

Then ψ is of class ∞, and $\psi \geqq 0$. Let

$$\phi_{cd}(x) = \int_c^x \psi_{cd}(t)\, dt \Big/ \int_c^d \psi_{cd}(t)\, dt.$$

Then ϕ is of class ∞, $0 \leqq \phi \leqq 1$, $\phi = 0$ for $x \leqq c$, and $\phi = 1$ for $x \geqq d$. If (a, b), (a', b') are two intervals and $a' < a$, $b < b'$, then, by piecing together two such functions as ϕ, we obtain a function

$$g(x) = \begin{cases} \phi_{a'a}(x) & \text{for } x \leqq b, \\ 1 - \phi_{bb'}(x) & \text{for } x > b \end{cases}$$

of class ∞, $0 \leqq g \leqq 1$, $g = 1$ in (a, b), $g = 0$ outside (a', b'). Let $g_i(x_i)$ be such a function for the pair (a_i, b_i), (a_i', b_i'). Then the product $g(x_1, \cdots, x_n) = g_1(x_1) \cdot \cdots \cdot g_n(x_n)$ has the properties asserted in (A).

(B). Let U be an open set in E^n with compact closure $\bar{U}$, and let U' be an open set containing $\bar{U}$. Then there exists a real-valued function g defined in E^n of class ∞, and such that

$$0 \leqq g \leqq 1, \qquad g = \begin{cases} 1 & \text{on } \bar{U}, \\ 0 & \text{outside } U'. \end{cases}$$

As $\bar{U}$ is compact, we can choose a finite number $D_1, \cdots, D_m$ of rectangular domains covering $\bar{U}$ such that the closure of each is in U'. Let D_i' be a rec-

[6a] A shorter proof can be based on a result of H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. Amer. Math. Soc., 36 (1934), pp. 63–89, Th. 3.

tangular domain containing $\bar{D}_i$ and contained in U'. Then there is a function g_i attached to the pair D_i, D_i' as in (A). Define the function g in E^n by

$$1 - g = (1 - g_1)(1 - g_2) \cdots (1 - g_m).$$

Then g is of class ∞, $0 \leqq g \leqq 1$, some $g_i = 1$ implies $g = 1$, every $g_i = 0$ implies $g = 0$. Thus $g = 1$ in $\sum D_i$, and $g = 0$ outside $\sum D_i'$.

(C). If Q is a point of M'', and $C(x)$ a coördinate neighborhood of $\pi(Q)$, then a coördinate neighborhood $C''(y)$ of Q in M'' may be chosen such that π has the form $x^i = \pi^i(y) \equiv y^i$ $(i = 1, \cdots, n)$.

Let $C(\bar{y})$ be a neighborhood of Q, and let π be given by the functions $x^i = \bar{\pi}^i(\bar{y})$. Since π has rank n at Q, the equations $y^i = \bar{\pi}^i(\bar{y})$ can be solved for n of the $\bar{y}$'s, say $\bar{y}_1, \cdots, \bar{y}_n$, as functions of the y's and the other $\bar{y}$'s in some neighborhood of Q. If we let $y^i = \bar{y}^i$ $(i = n + 1, \cdots, m)$, then, in this neighborhood, the y's form an admissible coördinate system.

Returning now to the proof proper of Theorem 2, let f be a map of M in M'' such that $\pi f =$ identity. Let $P \in M$ and $f(P) = Q \in M''$. Corresponding to a neighborhood C of P, let C'' be the neighborhood of Q given by (C). We can then choose two neighborhoods D, E of P such that $C \supset \bar{D}$, $D \supset \bar{E}$, and $f(\bar{D}) \subset C''$. As M is locally compact and separable, we may choose a sequence $\{C_i, C_i'', D_i, E_i\}$ of such sets of neighborhoods such that M is covered by the neighborhoods E_i, and each $\bar{E}_i$ meets only a finite number of other $\bar{E}_j$. This latter property can also be arranged for the D_i by reducing them in size without losing the property $D_i \supset \bar{E}_i$.

In the pair of neighborhoods D_1, C_1'', f is given by m functions $y^i = f^i(x)$. Due to the choice of coördinates in C_1'' and $\pi f =$ identity, we have $f^i(x) \equiv x^i$ $(i = 1, \cdots, n)$. Choose a pair of neighborhoods U', U such that $D_1 \supset \bar{U}'$, $U' \supset \bar{U}$ and $U \supset \bar{E}_1$. If $\epsilon > 0$ is given, we can choose functions $\phi^j(x)$ $(j = n + 1, \cdots, m)$ in D_1 of class ∞ and such that $|\phi^j - f^j| < \epsilon$. Let $g(x)$ be the function given by (B) for the pair U', U. Define

$$\psi^j = g\phi^j + (1 - g)f^j \qquad \text{in } D_1.$$

Then ψ^j is of class ∞ in U, and $\psi^j = f^j$ outside U'. If ϵ is sufficiently small, the functions $f^1, \cdots, f^n, \psi^{n+1}, \cdots, \psi^m$ define a map f_1 of D_1 in C_1''. If $f_1 = f$ outside D_1, then f_1 is a map of M in M'' such that $\pi f_1 =$ identity, and f_1 is differentiable to the class of M'' in an open set about $\bar{E}_1$. Since only a finite number of $\bar{D}_i$ meet $\bar{D}_1$, by restricting ϵ, we can insure that f_1 maps $\bar{D}_i$ in C_i'' for each i.

Suppose a map f_k of M in M'' is given such that $\pi f_k =$ identity, f_k maps each $\bar{D}_i$ in C_i'', and f_k is differentiable to the class of M'' in an open set W about $\sum_1^k \bar{E}_i$. Choose a pair U', U of open sets such that $U \supset \bar{E}_{k+1} - W \cdot \bar{E}_{k+1}$, $U' \supset \bar{U}$, $D_{k+1} \supset \bar{U}'$ and $\bar{U}' \cdot \sum_1^k \bar{E}_i = 0$. Apply now the construction of the preceding paragraph to each of the last $m - n$ components of f_k in the neighborhood D_{k+1}. The components ψ_k^j will not only be differentiable in U but likewise wherever the f_k^j are differentiable, namely, in $W \cdot D_{k+1}$. This leads to a new map f_{k+1} which is differentiable in an open set about $\sum_1^{k+1} \bar{E}_i$.

Since $f_{k+1} = f_k$ on $\sum_1^k \bar{E}_i$, the sequence of functions $\{f_k\}$ so constructed converges to a map f' of M in M'' such that $f' = f_k$ on $\sum_1^k \bar{E}_i$. Therefore f' is differentiable to the class of M'' everywhere and $\pi f' =$ identity.

13. Proof of Lemma 1

Consider the closed curve $A_1^{-1}A_2$ beginning and ending at Q. By assumption it is contractible into Q leaving its end points fixed. The two maps of F_P'' in F_Q'' are homotopic along this curve. We must cover the contraction of $A_1^{-1}A_2$ into Q by a contraction of the homotopy into F_Q''. Let E be a 2-cell, and ψ a map of E in M so that its boundary ∂E is mapped into $A_1^{-1}A_2$. Subdivide E into simplexes so fine that the image of each lies wholly within a coördinate neighborhood of M. Now $A_1^{-1}A_2$ can be contracted into Q over $\psi(E)$ in a finite number of steps, each consisting of deforming an arc of the curve over a simplex of E. Thus, each step occurs in a single coördinate neighborhood. Suppose then $\phi(t, \theta)$, $(a \leqq t \leqq b, 0 \leqq \theta \leqq 1)$ is a homotopy of the arc $\phi_0(t)$ into $\phi_1(t)$ leaving the end points fixed and occurring in a neighborhood $C(x)$. Let $h(y, t)$ be a homotopy of F_P'' along $\phi_0(t)$. Now the part of M'' over $C(x)$ is a product space $C \times F''$. Hence, $h(y, t)$ is given by a pair of functions $h'(y, t) = \phi_0(t)$ in C and $h''(y, t)$ in F''. Define $h'(y, t, \theta) = \phi(t, \theta)$ and $h''(y, t, \theta) = h''(y, t)$. The second pair h', h'' define in $C \times F''$ a deformation of $h(y, t)$ into a homotopy $h_1(y, t)$ along $\phi_1(t)$ without altering $h(y, a)$ and $h(y, b)$. This proves the lemma.

14. Proof of Theorem 4

The proofs of (a) to (e) parallel closely proofs to be found in a paper by Eilenberg [2]. The modifications are all of the same type. We shall be dealing with a simplex σ of M; C will be a neighborhood of σ, $C \times F''$ will be a representation of the part of M'' over C. In each case our problem will be to construct, extend, or deform homotopically a map f of σ or $\partial\sigma$. The function f has, locally, two components $\pi f =$ identity in C, and λf in F''. It will be seen that Eilenberg's arguments apply in each case to the component λf.

(a). Suppose f is defined on M^{h+1}. Then for any σ^{h+1}, $\lambda f(\sigma^{h+1})$ is a cell in F'' whose boundary is $\lambda f(\partial\sigma^{h+1})$. Therefore $c(f, \sigma^{h+1}) = 0$. Conversely, $c(f, \sigma^{h+1}) = 0$ means that λf can be extended to a map of σ^{h+1} in F''. If we extend πf over σ^{h+1} to be the identity, we obtain the extension of f.

(b). Let σ^{h+2} be arbitrary, and $K = |\sigma^{h+2}|$. Now λf maps K^h in F''. If F'' is taken to be the space Y of Eilenberg, then the chain $c^{h+1}(\lambda f)$ of Eilenberg is identical with $c^{h+1}(f)$ over K. As the former is known to be a cocycle on K, it follows that σ^{h+2} enters $\delta c^{h+1}(f)$ with coefficient zero. As σ^{h+2} is arbitrary, the proof is complete.

(c). We need the following:

LEMMA 2. *There exists a homotopy $f(x, t)$ of $f(x)$ for $x \in M^h$, $0 \leqq t \leqq 1$, such that $\pi f(x, t) = x$ for $x \in M^h$ and all t, $f(x, 0) = f(x)$ for $x \in M^h$, and $f(x, 1) = f'(x)$ for $x \in M^{h-1}$.*

Define $f(x, 0) = f(x)$ for $x \in M^h$, $f(x, 1) = f'(x)$ for $x \in M^{h-1}$ and $\pi f(x, t) = x$ for $x \in M^h$, $0 \leqq t \leqq 1$. We must extend the local components $\lambda f(x, t)$ over the rest of $M^h \times I$. Order the simplexes of M^{h-1} in a sequence such that each simplex is preceded by its faces. As the first is a vertex P, and F'' is connected (if $h > 0$), we can join $f(P)$ to $f'(P)$ by a curve $f(P, t)$ in F''. Suppose $f(x, t)$ is defined over all prisms $\sigma \times I$ for simplexes σ preceding σ_i. Then $f(x, t)$ is defined on $\sigma_i \times 0$, $\sigma_i \times 1$ and $\partial\sigma_i \times I$, and therefore $\lambda f(x, t)$ maps $\partial(\sigma_i \times I)$ in F''. As the dimension of this sphere is $< h$, λf can be extended over $\sigma_i \times I$. Having thus defined $f(x, t)$ over $M^{h-1} \times I$, we have only to extend it over each $\sigma^h \times I$.

As $\lambda f(x, t)$ is defined over $\sigma^h \times 0 + \partial\sigma^h \times I$, and this set is a retract of $\sigma^h \times I$, $\lambda f(x, t)$ admits a continuous extension over $\sigma^h \times I$. This proves the lemma.

Let $f''(x) = f(x, 1)$. Since for any σ^{h+1}, $\lambda f(\partial\sigma^{h+1})$ is homotopic in F'' to $\lambda f''(\partial\sigma^{h+1})$, we have $c^{h+1}(f) = c^{h+1}(f'')$. Corresponding to the maps f', f'' which agree on M^{h-1}, we define (following Eilenberg) an h-chain $d^h(f', f'')$. Let σ^h be any h-simplex. Write the oriented h-sphere as a sum $S^h = E' + E''$ of two hemispheres. Let g', g'' map E', E'' respectively on σ^h topologically with degrees 1, -1 so that they agree on the common boundary. Then define $\phi = f'g'$ on E', $\phi = f''g''$ on E''. As f', f'' agree on $\partial\sigma^h$, $\lambda\phi$ is a continuous map of S^h in F'' determining thereby an element of $\pi_h(F'')$ which is denoted by $d(f', f'', \sigma^h)$. Let $d^h(f', f'') = \sum d(f', f'', \sigma^h)\sigma^h$. Then

$$\delta d^h(f', f'') = c^{h+1}(f') - c^{h+1}(f''). \tag{A}$$

If, for any σ^{h+1}, we consider the functions $\lambda f'$, $\lambda f''$ defined on $|\partial\sigma^{h+1}|$ and the chains $d^h(\lambda f', \lambda f'')$, $c^{h+1}(\lambda f')$, $c^{h+1}(\lambda f'')$ as defined by Eilenberg for the complex $|\sigma^{h+1}|$, then the relation analogous to (A) holds. However these latter chains agree with the former where both are defined. The proof of (c) is therefore complete.

(d). Let $\delta d^h = c^{h+1} - c^{h+1}(f)$. Define $f' = f$ on M^{h-1}. Given any σ^h, define $\pi f' =$ identity in σ^h. Then define $\lambda f'$ on σ^h so that $d(f', f, \sigma^h)$ (see proof of (c)) is the coefficient of σ^h in d^h. Then $d^h(f', f) = d^h$. As shown in (c), $\delta d^h(f', f) = c^{h+1}(f') - c^{h+1}(f)$. Therefore, $c^{h+1} = c^{h+1}(f')$.

(e). If a map f of M^{h+1} exists, then, by (a), $c^{h+1}(f) = 0$. By (c), for any f' defined on M^h, $c^{h+1}(f') \backsim c^{h+1}(f) = 0$. Suppose $c^{h+1}(f') \backsim 0$ for some map of M. Then, if we choose $c^{h+1} = 0$ in (d), we obtain a new map f'' of M^h such that $c^{h+1}(f'') = 0$. It follows from (a) that f'' can be extended over M^{h+1}.

(f). Let M_1, M_2 be two subdivisions of M into complexes. We suppose M_2 is so fine that there is a simplicial map τ of M_2 into M_1 such that, for each point x, $\tau(x)$ lies on the closure of the simplex of M containing x. There is, therefore, a homotopy $\tau(x, t)$ connecting the identity to $\tau(x)$ such that $\tau(x, t)$ lies on the closure of the simplex of M_1 containing x for $0 \leqq t \leqq 1$. We now suppose M_1 is so fine that the map τ of any prism $\zeta \times I$, where ζ is a simplex of M_2, lies in a neighborhood of M. (This will be the case if the closure of the star of each simplex of M_1 lies in a neighborhood). Now let f be a map of M_1^h in M''. Construct a map f' of M_2^{h-1} in M'' (always so that $\pi f' =$ identity). We must extend

f' over M_2^h so that $c_2^{h+1}(f') \frown c_1^{h+1}(f)$. To this end we shall define a map $f'(x, t)$ of $M_2^h \times I$ in M''. Let $f'(x, 1) = f\tau(x)$ for $x \in M_2^h$, and $f'(x, 0) = f'(x)$ for $x \in M_2^{h-1}$. Order the simplexes of M_2^h in a sequence so that each is preceded by its faces. If P is the first (a vertex), we define $\pi f'(P, t) = \tau(P, t)$ and $\lambda f'(P, t)$ to be a path in F'' joining $\lambda f'(P, 0)$ to $\lambda f'(P, 1)$. Suppose now $f'(x, t)$ is defined for x in any simplex preceding ζ, and is such that $\pi f'(x, t) = \tau(x, t)$. Then, if dimension $\zeta < h$, $f'(x, t)$ is defined on the boundary of $\zeta \times I$, and therefore, $\lambda f'(x, t)$ admits an extension over this prism. We extend $\pi f'(x, t)$ so as to be $\tau(x, t)$ over $\zeta \times I$. If, however, dimension $\zeta = h$, $f'(x, t)$ is defined only on $\partial\zeta \times I + \zeta \times 1$. As this set is a retract of $\zeta \times I$, $\lambda f'(x, t)$ admits an extension mapping $\zeta \times I$ in F''. Then we set $\pi f'(x, t) = \tau(x, t)$. Now let $f'(x) = f'(x, 0)$ for $x \in M_2^h$. Then, for any ζ^{h+1} of M_2^h, $\lambda f'(\partial\zeta^{h+1})$ is homotopic in F'' to $\lambda f(\partial\tau\zeta^{h+1})$. Therefore $c_2(f', \zeta^{h+1}) = c_1(f, \tau\zeta^{h+1})$. Therefore, the cocycle $c_2^{h+1}(f')$ is the image of the cocycle $c_1^{h+1}(f)$ under the simplicial map τ. This completes the proof of (f).

Appendix I. Extensions of method

In some problems a tensor may already be defined on a submanifold L of M, and it may be required to extend it over M. Such a problem could arise in connection with a manifold M with regular boundary L. If M is subdivided so that L is a subcomplex, then as before f extends to $L + M^h$. The chain $c^{h+1}(f)$ lies in $M - L$ and is a cocycle in the open complex $M - L$. Then a necessary and sufficient condition that f on L can be extended to $L + M^{h+1}$ is that $c^{h+1}(f) \sim 0$ in $M - L$.

In those problems for which the characteristic cohomology class is zero and maps f of M^{h+1} can be defined, we are faced with the difficulty of extending such an f to M^{h+2}. We can, of course, proceed as before and define a chain $c^{h+2}(f)$ with coefficients in $\pi_{h+1}(F'')$. For this, it is necessary to know that F'' is $(h + 1)$-simple in the sense of Eilenberg,[7] so that $\lambda f(\partial\sigma^{h+2})$ determines a unique element of $\pi_{h+1}(F'')$. It will follow as before that $c^{h+2}(f)$ is a cocycle; and $c^{h+2}(f) \sim 0$ is a sufficient condition for an f' defined on M^{h+2} to exist. However, it is not necessary; for the cohomology class of $c^{h+2}(f)$ may well vary from one f to another. Just how this cohomology class varies is not known. Some deeper analysis is required to resolve the difficulty.

Appendix II. Application to fibre spaces

In a paper by Hurewicz and the author [3], the notion of a fibre space was introduced as a generalization of the Whitney notion of fibre bundle. In those cases for which we know that F'' is compact, we can prove that M'' is a fibre space over M relative to π. This suggests that any fibre space X over a base space B relative to a mapping $\pi(X) = B$ determines in some sense a characteristic cohomology class in B. We propose to show that this is the case.

Assuming B to be arcwise connected, the fibres F may be deformed along

[7] If $h > 1$, then $\pi_1(F'') = 0$, and F'' is i-simple for every i.

curves into others and these deformations induce isomorphisms between the $\pi_h(F)$ (where h is the smallest integer such that $\pi_h(F) \neq 0$). Therefore, the coefficient groups $\pi_h(F)$ are at hand as before.

A fibre space X need not be locally a product space over its base space B. Consequently, the πf, λf method must be replaced by a new mechanism. Such a one is provided by the following lemma.

LEMMA 3. *Let X be a fibre space over B relative to π. Let K be a complex, L a subcomplex, ψ a map of K in B, and ψ' a map of L in X such that $\pi\psi' = \psi$. Suppose ψ_t is a homotopy of $\psi(K)$ and ψ_t' is a homotopy of $\psi'(L)$ such that $\pi\psi_t' = \psi_t$, $0 \leqq t \leqq 1$. Then, if ψ_1' admits an extension to K such that $\pi\psi_1' = \psi_1$, this is also true of ψ'.*

The complex $U = L\times I + K\times 1$ is a deformation retract of $K\times I$. Let h_τ $(0 \leqq \tau \leqq 1)$ be such a retraction. Since ψ_t' gives a map of U in X such that $\pi\psi_t' = \psi_t$ on U, the function $\psi_t' h_1$ maps $K\times I$ in X such that $\pi\psi_t' h_1 = \psi_t h_1$. Since $\psi_t h_1$ is homotopic to ψ_t leaving U pointwise fixed, the covering homotopy [3] deforms $\psi_t' h_1$ leaving U fixed into an extension of ψ_t' to $K\times I$ such that $\pi\psi_t' = \psi_t$. Then ψ_0' is the required extension of ψ'.

Returning now to the problem of defining a characteristic cocycle, let K be a complex and ψ a map of K in B. We propose to show there is a map ψ' of K^h in X such that $\pi\psi' = \psi$. There is no difficulty in defining ψ' on K^0. Suppose it has been properly defined on K^i and σ is an $(i+1)$-simplex. Then ψ is defined on σ, and ψ' on $\partial\sigma$. Let h_t contract σ on itself to a point. Let $\psi_t = \psi h_t$, and let ψ_t' cover ψ_t on $\partial\sigma$. If $i < h$, the map ψ_1' of $\partial\sigma$ in the fibre F over $\psi_1(\sigma)$ can be extended to a map of σ in F. Applying Lemma 3 gives an extension of ψ' over σ. Thus, ψ' can be defined over K^h.

If ψ' is a map of K^h in X such that $\pi\psi' = \psi$, we can define a chain $c^{h+1}(\psi')$ in K as follows. Let σ be an $(h+1)$-simplex, and let h_t contract $\partial\sigma$ over σ to a point P_0. Then the homotopy $\psi_t = \psi h_t$ of $\partial\sigma$ has a covering homotopy ψ_t' of ψ'. As ψ_1' maps $\partial\sigma$ into the fibre F over $\psi(P_0)$, it defines an element of $\pi_h(F)$ which we denote by $c(\psi', \sigma^{h+1})$. Define $c^{h+1}(\psi') = \sum c(\psi', \sigma^{h+1})\sigma^{h+1}$. Then we have:

THEOREM 5. *If X is a fibre space over B relative to π, K a complex, and ψ a map of K in B, then there exist maps ψ' of K^h in X such that $\pi\psi' = \psi$, where h is the smallest integer for which $\pi_h(F) \neq 0$. Any such map determines a cocycle $c^{h+1}(\psi')$ in K whose cohomology class c^{h+1} is independent of ψ'. It may be chosen arbitrarily in its class by an appropriate choice of ψ'. Therefore, in order that there be a map ψ' of K^{h+1} in X such that $\pi\psi' = \psi$, it is necessary and sufficient that the cohomology class $c^{h+1} = 0$. Furthermore, the class c^{h+1} is independent of the subdivision used in K.*

The proofs are like those of Theorem 4. They differ only in that covering homotopies and Lemma 3 are used in place of the $C\times F''$ construction.

A continuous finite cycle in B is composed of three things: a complex K, a map ψ of K in B, and a finite cycle Z in K. Let us consider $(h+1)$-cycles with local coefficients in the character groups of the $\pi_h(F)$. Then $c^{h+1}\cdot Z^{h+1}$ is a real

number mod 1. If (K_1, ψ_1, Z_1) and (K_2, ψ_2, Z_2) are homologous continuous cycles, then by definition, there is a complex K_3 containing K_1, K_2 as subcomplexes, a map ψ_3 of K_3 in B agreeing with ψ_1, ψ_2 on K_1, K_2, and a chain in K_3 whose boundary is $Z_1 - Z_2$. If a map ψ_3' of K_3^h in X is given such that $\pi\psi_3' = \psi_3$, then $c_1^{h+1}(\psi_3')$, $c_2^{h+1}(\psi_3')$ and $c_3^{h+1}(\psi_3')$ are simultaneously defined, $c_1^{h+1}(\psi_3')$ is the part of $c_3^{h+1}(\psi_3')$ on K_1, and $c_2^{h+1}(\psi_3')$ the part on K_2. Therefore, $c_1^{h+1}\cdot Z_1 = c_3^{h+1}\cdot Z_1$, $c_2^{h+1}\cdot Z_2 = c_3^{h+1}\cdot Z_2$. Since $Z_1 \sim Z_2$ in K_3, $c_3^{h+1}\cdot Z_1 = c_3^{h+1}\cdot Z_2$. Thus, the real number mod 1, $c^{h+1}\cdot Z$, is independent of the representation of the homology class of Z. Thus, a map of the homology group $H^{h+1}(B)$ in the reals mod 1 is at hand. This map is homomorphic; for the sum of two cycles is represented by the abstract sum $K_1 + K_2$, a map $\psi = \psi_1$ on K_1, $= \psi_2$ on K_2, and the sum $Z_1 + Z_2$. The characteristic cocycle on $K_1 + K_2$ is the sum of the cocycles of K_1 and of K_2. Therefore, $c^{h+1}\cdot(Z_1 + Z_2) = c_1^{h+1}\cdot Z_1 + c_2^{h+1}\cdot Z_2$. Summarizing, we have:

THEOREM 6. *Corresponding to the fibre space X over B, there is a characteristic cohomology class c^{h+1} in B, where h is the smallest integer such that $\pi_h(F) \neq 0$. c^{h+1} belongs to the cohomology group of characters of the homology group $H^{h+1}(B)$ based on continuous finite cycles with local coefficients in the character groups of the $\pi_h(F)$. If K is a complex, and ψ a map of K in B, then the image of c^{h+1} in K under ψ is the characteristic cohomology class of K relative to ψ (see Theorem 5).*

THE UNIVERSITY OF CHICAGO

BIBLIOGRAPHY

1. S. S. CAIRNS, *Triangulation of the manifold of class one*, Bull. Amer. Math. Soc., 41 (1935), pp. 549–552.
2. S. EILENBERG, *Cohomology and continuous mappings*, Annals of Math., 41 (1940), pp. 231–251.
3. W. HUREWICZ AND N. E. STEENROD, *Homotopy relations in fibre spaces*, Proc. Nat. Acad. Sci., 27 (1941), pp. 61–64.
4. W. HUREWICZ, *Beiträge zur Topologie der Deformationen* I–IV, Proc. Amsterdam Acad., 38 (1935), pp. 112 and 521, also 39 (1936), pp. 117 and 215.
5. E. STIEFEL, *Richtungsfelder und Fernparallelismus in n-dimensionalen Mannigfaltigkeiten*, Comm. Math. Helv., 8 (1936), pp. 3–51.
6. H. WHITNEY, *On the theory of sphere-bundles*, Proc. Nat. Acad. Sci., 26 (1940), pp. 148–153.
7. H. WHITNEY, *Differentiable manifolds*, Annals of Math., 37 (1936), pp. 645–680.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

HOMOTOPY PROPERTIES OF THE REAL ORTHOGONAL GROUPS

BY GEORGE W. WHITEHEAD

(Received September 25, 1941)

1. Introduction

In this paper we propose to investigate the topological structure of the rotation group R_n of the n-sphere, with special emphasis on its homotopy properties. Of particular interest are the homotopy groups π_i of R_n. These groups, one for each dimension i, were first defined for a general space by Hurewicz [1].[1] Like the homology groups, they are topological invariants of a space; unlike the homology groups, however, no general method for computing them is known. Each space thus presents a problem in itself.

The computation of the groups $\pi_i(R_n)$ for $i \leqq 5$ and all n will be carried out by the method of fibre mappings and covering homotopies developed by Hurewicz and Steenrod [2]. We shall make extensive use of the results of Freudenthal [3], Hopf [4], and Pontrjagin [5] on the homotopy groups of spheres.

The groups $\pi_i(R_n)$ are useful not only in the study of the homotopy properties of spheres, but also are used by Whitney in his theory of sphere-bundles [6, 7], where they appear as coefficient groups for certain cocycle invariants.

Another application of our results appears in the theory of continuous vector fields over spheres. It is well known that no continuous field of unit vectors can be defined over the spheres of even dimension. Over the odd-dimensional spheres, however, one such vector field can always be defined; and if $n \equiv 3 \pmod 4$ or $n \equiv 7 \pmod 8$ it is possible to define three or seven independent vector fields, respectively, over the n-sphere S^n. These can be readily constructed by the use of the multiplication matrices for quaternions and Cayley numbers. For a general odd n, however, there is no known result on the maximum number of independent vector fields which can be defined over S^n.

In this paper the case $n \equiv 1 \pmod 4$ is resolved as follows: *Any two vector fields over* S^{4m+1} $(m = 0, 1, 2, \cdots)$ *are somewhere dependent.* As a corollary to this result it is observed that *the tangent sphere-bundle of* S^n *is not simple if* $n > 1$ *and* $n \equiv 1 \pmod 4$.

This investigation was carried out under the direction of Prof. N. E. Steenrod, to whom the author wishes to acknowledge his indebtedness for many valuable suggestions and criticisms.

2. Table of groups π_1 to π_5 of R_n

's section the results of our computation of the homotopy groups $\pi_i(R_n)$ are exhibited, and a set of generators for these groups is given. Proofs will be deferred until Section 8.

[1] Numbers in square brackets refer to the bibliography at the end of the paper.

Let ∞ denote the free cyclic group, 2 the cyclic group of order two. If A and B are two abelian groups, $A + B$ denotes their direct sum. In terms of these notations, the groups $\pi_i(R_n)$ may be tabulated as follows:

	R_1	R_2	R_3	R_4	R_5	R_6	$\cdots$	R_n	$\cdots$
π_1	∞	2	2	2	2	2	$\cdots$	2	$\cdots$
π_2	0	0	0	0	0	0	$\cdots$	0	$\cdots$
π_3	0	∞	$\infty + \infty$	∞	∞	∞	$\cdots$	∞	$\cdots$
π_4	0	2	$2 + 2$	2	0	0	$\cdots$	0	$\cdots$
π_5	0	0	0	0	∞	0	$\cdots$	0	$\cdots$

The results of the first two rows were first obtained by Cartan [8].

A generator of $\pi_1(R_n)$ is given by the map of the circle $x_1^2 + x_2^2 = 1$ defined by

$$x \to \begin{Vmatrix} x_1 & -x_2 & 0 \\ x_2 & x_1 & 0 \\ 0 & 0 & I_{n-1} \end{Vmatrix},$$

where I_{n-1} is the $(n-1)$-rowed identity matrix.

A generator of $\pi_3(R_n)$ $(n \geqq 3)$ is given by the map of the 3-sphere $\sum_{i=1}^{4} x_i^2 = 1$

$$x \to \begin{Vmatrix} x_1 & -x_2 & -x_3 & -x_4 & 0 \\ x_2 & x_1 & -x_4 & x_3 & 0 \\ x_3 & x_4 & x_1 & -x_2 & 0 \\ x_4 & -x_3 & x_2 & x_1 & 0 \\ 0 & 0 & 0 & 0 & I_{n-3} \end{Vmatrix}.$$

The corner matrix is the linear transformation of Euclidean 4-space defined by multiplying every quaternion on the left by $x_1 + ix_2 + jx_3 + kx_4$.

The generator of $\pi_3(R_2)$ is the well-known double covering of R_2 by S^3. Bordering this matrix with a 1 in the lower right hand corner and zeros elsewhere, we obtain the extra generator of $\pi_3(R_3)$.

To obtain the generator of $\pi_4(R_2)$, we map S^4 on S^3 essentially, and then map S^3 into R_2 by means of the generator of $\pi_3(R_2)$ given above. Such an essential map was constructed by Freudenthal [3]. In a similar manner we obtain generators for $\pi_4(R_3)$ and $\pi_4(R_4)$.

Finally, the generator of $\pi_5(R_5)$ is determined by the map of S^5 into R_5 given by

$$x \to \| \delta_{ij} - 2x_i x_j \| \cdot \begin{Vmatrix} I_5 & 0 \\ 0 & -1 \end{Vmatrix}.$$

3. Preliminary notions

In this section we introduce notations and concepts which will occur throughout the paper. The relative and absolute homotopy groups of a space are introduced and two homomorphisms relating these groups are discussed.

Let points x in Euclidean $(n+1)$-space be referred to coördinates $(x_1, x_2, \cdots, x_{n+1})$. The unit sphere $\sum x_i^2 = 1$ we denote by S^n. The *equatorial plane* $x_{n+1} = 0$ divides S^n into two hemispheres V_1^n and V_2^n defined by the inequalities $x_{n+1} \geqq 0$ and $x_{n+1} \leqq 0$ respectively. If $x = (x_1, x_2, \cdots, x_{n+1}) \in S^n$, the *antipodal point* $(-x_1, -x_2, \cdots, -x_{n+1})$ is denoted by $\bar{x}$. We shall refer to the point $x^o = (0, 0, \cdots, 1)$ as the *north pole*, and to its antipode $\bar{x}^o$ as the *south pole*.

The group R_n of all rotations of S^n may be represented as the group of all real square orthogonal matrices of order $n+1$ with determinant $+1$. The subgroup of R_n consisting of all those rotations of S^n which leave the north pole fixed is isomorphic with the group R_{n-1}, and we shall denote the former group also by the symbol R_{n-1}.

Let Y be a topological space, F a closed subset of Y, and y_o a fixed point of F. Let $\mathfrak{X}$ denote the space of all maps of V_1^n into Y which carry the boundary $\partial V_1^n = S^{n-1}$ into F and the north pole x^o of S^{n-1} into y_o. We introduce an equivalence relation in $\mathfrak{X}$ as follows: two maps $f_1, f_2 \in \mathfrak{X}$ are said to be *equivalent* if they are homotopic, and during the homotopy S^{n-1} remains in F and x^o remains at y_o. In other words, two points of $\mathfrak{X}$ are equivalent if they can be joined by an arc in $\mathfrak{X}$. The relation of equivalence is easily seen to be reflexive, symmetric, and transitive, and thus divides $\mathfrak{X}$ into classes of equivalent maps, called *homotopy classes*. The homotopy class determined by a map f we denote by $\{f\}$; the set of all such homotopy classes by $\pi_n(Y, F)$. Hurewicz [1] has introduced an operation, called addition, in $\pi_n(Y, F)$, by means of which it becomes a group, the n^{th} *relative homotopy group of* Y *modulo* F. If the closed set F is specialized to consist only of the point y_o, the group $\pi_n(Y, y_o)$ so obtained is called the *absolute homotopy group* $\pi_n(Y)$. The latter group may also be defined by means of mappings of spheres into Y and we shall frequently find it convenient to use this definition.

We now introduce a homomorphism ω of $\pi_n(Y, F)$ into $\pi_{n-1}(F)$. This homomorphism is defined as follows: if $\alpha = \{f\} \in \pi_n(Y, F)$, then $\omega(\alpha)$ denotes the homotopy class of $\pi_{n-1}(F)$ determined by the map $f(S^{n-1}) \subset F$. Evidently $\{f\} = \{g\}$ implies $\omega(\{f\}) = \omega(\{g\})$. Thus ω maps $\pi_n(Y, F)$ into $\pi_{n-1}(F)$, and it follows from the definition of addition that ω is a homomorphism. Let $\pi_{no}(Y, F)$ denote the kernel of this homomorphism, $\pi_{n-1,o}(F)$ the image of $\pi_n(Y, F)$ under ω. Evidently $\pi_{no}(Y, F)$ consists of those homotopy classes determined by those relative n-cells in Y modulo F which are contractible into F, while $\pi_{n-1,o}(F)$ consists of the classes determined by those $(n-1)$-spheres in F which are homotopic to points in Y.

Since each element of $\pi_n(Y)$, considered as a set, is a subset of an element of $\pi_n(Y, F)$, we have a natural mapping ψ of $\pi_n(Y)$ into $\pi_n(Y, F)$, which is a homomorphism. It is not hard to show that $\psi(\{f\}) = 0$ if and only if some f' in the

class of f maps S^n into F. If $\{f_1\} = \{f_2\}$ and $\psi(\{f_1\}) = \psi(\{f_2\}) = 0$, then f_1' is homotopic in Y to f_2'. Thus f_1' and f_2' determine the same element of $\pi_n(F)/\pi_{no}(F)$. Conversely, if $f(S^n) \subset F$, then $\psi(\{f\}) = 0$. Hence the kernel of the homomorphism ψ is isomorphic to $\pi_n(F)/\pi_{no}(F)$. The image of $\pi_n(Y)$ under ψ is evidently the group $\pi_{no}(Y, F)$. We summarize these results in

THEOREM 1. *The natural homomorphic maps $\omega[\pi_n(Y, F)] \subset \pi_{n-1}(F)$ and $\psi[\pi_n(Y)] \subset \pi_n(Y, F)$ are related as follows: the kernel of the homomorphism ψ is isomorphic to $\pi_n(F)/\pi_{no}(F)$, while the image of $\pi_n(Y)$ under ψ is the group $\pi_{n,o}(Y,F)$.*

4. Homotopy relations in compact Lie groups

Let G be a topological group, H a closed subgroup of G, and $B = G/H$ the space of left (or right) cosets of H in G. Then there is a natural mapping π of G onto B defined as follows: for every $g \in G$, $\pi(g)$ is the coset of B containing g.

If G is a compact Lie group, then π is a fibre map in the sense of Hurewicz and Steenrod [2]. A slicing function can be defined as follows: a plane of maximum dimension independent of the tangent plane to H at the identity 1 meets each coset b in a sufficiently small neighborhood U of $b_o = \pi(1)$ just once. We denote this point by $\phi(1, b)$. Then if $g^{-1}b \in U$, let $\phi(g, b) = g\phi(1, g^{-1}b)$. Evidently ϕ has all the required properties of a slicing function.

The following theorem will be useful in our discussion of the rotation groups:

THEOREM 2. *If G is a topological group and B is the space of left (or right) cosets of a closed subgroup H of G, and if there exists a map $f(B) \subset G$ such that $\pi f(b) = b$, then G is homeomorphic with the product space $H \times B$.*

We may suppose B is a space of left cosets. Let $f'(b) = f(b)\cdot[f(b_o)]^{-1}$; then $\pi f'(b) = \pi f(b) = b$, and $f'(b_o) = 1$. We then set up the homeomorphism by means of two maps $p(G) = H \times B$ and $q(H \times B) = G$ defined as follows:

$$p(g) = [g^{-1}\cdot f'(\pi g), \pi g] \qquad (g \in G),$$

$$q(h, b) = f'(b)\cdot h^{-1} \qquad (h \in H, b \in B).$$

Then

$$p[q(h, b)] = \{h[f'(b)]^{-1}f'(b), b\} = (h, b),$$

$$q[p(g)] = f'(\pi g)\cdot[f'(\pi g)]^{-1}\cdot g = g,$$

and both maps are continuous. Hence G and $H \times B$ are homeomorphic.

5. Slicing functions for $R_n \to R_n/R_{n-1}$

In this section the results of Section 4 are applied to the special case $G = R_n$, $H = R_{n-1}$, and an explicit slicing function is constructed. The special cases $n = 1, 3, 7$ are treated separately, and for these values of n Theorem 2 is applied.

Let us consider the mapping $\pi(R_n) = S^n$ defined by $\pi(r) = r(x^o)$ $(r \in R_n)$. As shown by Hurewicz and Steenrod [2], π is a fibre map of R_n into S^n, the fibres being the left cosets of R_{n-1} in R_n. In terms of the matrix representation of R_n, $\pi(r)$ is the last column of the matrix r.

We now define the slicing function promised above: if $x \neq \tilde{x}^o$, let $\phi(I, x)$ be the rotation carrying x^o along a great circle into x, and leaving the $(n-2)$-sphere orthogonal to this great circle fixed. In terms of matrices

$$\phi(I, x) = A_n(x) = \left\| \delta_{ij} - \frac{(x_i + \delta_{i,n+1})(x_j + \delta_{j,n+1})}{x_{n+1} + 1} \right\| \cdot \begin{Vmatrix} I_n & 0 \\ 0 & -1 \end{Vmatrix} \tag{1}$$

$$(i, j = 1, \cdots, n+1),$$

where I_n denotes the n-rowed identity matrix. Then $\phi(r, x)$ is defined as is Section 2. Evidently it is impossible to extend $\phi(I, x)$ so as to be defined and continuous over all of S^n. We shall show later that there is no slicing function with this property for a general n.

For the dimensions 1, 3, and 7, however, it is possible to define such a slicing function. Let $\mathfrak{A}_{n+1}$ $(n = 1, 3, 7)$ denote the algebras of complex numbers, quaternions, and Cayley numbers, respectively, over the field of real numbers. By means of these algebras a multiplication $x \cdot y$ of points of Euclidean $(n+1)$-space is defined. This multiplication has the property that $\| x \cdot y \| = \| x \| \cdot \| y \|$, where $\| x \|^2 = \sum x_i^2$ is the square of the distance of the point x from the origin. Hence S^n is closed under multiplication and for $x, y \in S^n$ we have $x \cdot y = B_n(x) \cdot y$, where $B_n(x) \in R_n$, and, if the coördinate system is chosen so that x^o is the unit of the algebra, $B_n(x^o) = I$, $\pi B_n(x) = x$. Since $B_n(x)$ is defined for all $x \in S^n$, we have

THEOREM 3. *For $n = 1, 3, 7$, R_n is a product space $R_{n-1} \times S^n$.*

Since the i^{th} homotopy group of a product space $X \times Y$ is the direct sum of the i^{th} homotopy groups of X and Y, we have

COROLLARY. *For $n = 1, 3, 7$, $\pi_i(R_n)$ is the direct sum $\pi_i(R_{n-1}) + \pi_i(S^n)$; in particular, $\pi_{n-1}(R_{n-1}) = \pi_{n-1}(R_n)$.*

6. The canonical map of S^n in R_n

We now introduce a mapping C_n of S^n into R_n which plays an important role in the following discussion. We shall refer to it as the *canonical map*. It is proved that this map is contractible into R_{n-1} if n is even; while for n odd it is not so contractible. The canonical map is then used to construct a generator for $\pi_n(R_n, R_{n-1})$.

In order to define the map C_n, let $\theta(x)(x \in S^n)$ denote the angular distance from x^o to x, and let x' be the point in the great circle through x^o and x with $\theta(x') = 2\theta(x)$. Let $C_n(x)$ $(x \neq \tilde{x}^o)$ be the rotation which carries x^o along a great circle into x' and leaves the orthogonal $(n-2)$-sphere fixed; and let $C_n(\tilde{x}^o) = I$. Evidently

$$C_n(x) = [A_n(x)]^2 = \| \delta_{ij} - 2x_i x_j \| \cdot \begin{Vmatrix} I_n & 0 \\ 0 & -1 \end{Vmatrix} \qquad (i, j = 1, \cdots, n+1). \tag{2}$$

We observe that C_n, unlike A_n, is defined and continuous over all of S^n. Furthermore, antipodal points have the same image, while distinct pairs of antipodal points have distinct images. Thus the image of S^n in R_n is a projective n-space.

Let $g_n(S^n) = S^n$ denote the projection πC_n of the canonical map. If $g_{n,i}(x)$ is the i^{th} coördinate of $g(x)$, we have $g_{n,i}(x) = 2x_i x_{n+1} - \delta_{i,n+1}$ $(i = 1, \cdots, n+1)$.

THEOREM 4. *If n is even, g_n has degree zero; if n is odd, g_n has degree two.*

For g_n maps the equator S^{n-1} into $\tilde{x}^o$ and maps $V_1^n - S^{n-1}$ topologically on $S^n - \tilde{x}^o$; in fact, g_n can be obtained by a homotopy of S^n on itself in which each point moves along the great circle joining it to the north pole. Thus g_n maps V_1^n on S^n with degree 1. Furthermore, g_n maps V_2^n on S^n with degree $(-1)^{n+1}$; for $g_n(x) = g_n(\tilde{x})$ $(x \in V_2^n)$ and the antipodal transformation $x \to \tilde{x}$ has degree $(-1)^{n+1}$. Hence g_n maps S^n on itself with degree $1 + (-1)^{n+1}$.

If n is even, g_n has degree zero, and hence is homotopic to a point. We shall give a homotopy of g_n which will be useful in a later section. This homotopy $g_n(x, t)$ is given by the equations

$$
\begin{aligned}
g_{n,2i-1}(x, t) &= 2\{(1-t)x_{2i-1}x_{n+1} + [t(1-t)]^{\frac{1}{2}}x_{2i}\},\\
g_{n,2i}(x, t) &= 2\{(1-t)x_{2i}x_{n+1} - [t(1-t)]^{\frac{1}{2}}x_{2i-1}\} \qquad (i = 1, \cdots, n/2),\\
g_{n,n+1}(x, t) &= 1 - 2(1-t)(1-x_{n+1}^2).
\end{aligned} \tag{3}
$$

It is easy to verify that $g_n(x, t)$ contracts $g_n(S^n)$ over S^n into x^o.

If n is odd, g_n has degree two. We shall give a deformation of g_n into a second map g_n' of degree two. The latter map is defined by the equations

$$
\begin{aligned}
g_{n,i}'(x) &= x_i \qquad (i = 1, \cdots, n-1);\\
g_{n,n}'(x) &= \frac{2x_n x_{n+1}}{(x_n^2 + x_{n+1}^2)^{\frac{1}{2}}},\\
g_{n,n+1}'(x) &= \frac{x_{n+1}^2 - x_n^2}{(x_n^2 + x_{n+1}^2)^{\frac{1}{2}}} \qquad (x_n^2 + x_{n+1}^2 \neq 0),\\
g_{n,n}'(x) &= g_{n,n+1}'(x) = 0 \qquad (x_n = x_{n+1} = 0).
\end{aligned} \tag{4}
$$

It is easily seen that g_n' is defined and continuous over all of S^n. The homotopy $g_n(x, t)$ of g_n over S^n into g_n' is given by

$$
\begin{aligned}
g_{n,2i-1}(x, t) &= tx_{2i-1} + [t(1-t)]^{\frac{1}{2}}x_{2i} + 2x_{n+1}\frac{(1-t)x_{2i-1} - [t(1-t)]^{\frac{1}{2}}x_{2i}}{\{1 - t[1 - (x_n^2 + x_{n+1}^2)]\}^{\frac{1}{2}}},\\
g_{n,2i}(x, t) &= tx_{2i} - [t(1-t)]^{\frac{1}{2}}x_{2i-1} + 2x_{n+1}\frac{(1-t)x_{2i} + [t(1-t)]^{\frac{1}{2}}x_{2i-1}}{\{1 - t[1 - (x_n^2 + x_{n+1}^2)]\}^{\frac{1}{2}}}\\
&\qquad \left(i = 1, \cdots, \frac{n-1}{2}\right),
\end{aligned} \tag{5}
$$

$$
\begin{aligned}
g_{n,n}(x, t) &= \frac{2x_n x_{n+1}}{\{1 - t[1 - (x_n^2 + x_{n+1}^2)]\}^{\frac{1}{2}}},\\
g_{n,n+1}(x, t) &= \frac{2x_{n+1}^2}{\{1 - t[1 - (x_n^2 + x_{n+1}^2)]\}^{\frac{1}{2}}} - \{1 - t[1 - (x_n^2 + x_{n+1}^2)]\}^{\frac{1}{2}}.
\end{aligned}
$$

Although $g_n(x, t)$ is not defined everywhere for $t = 1$, it is easily verified that $\lim_{t\to 1} g_n(x, t) = g'_n(x)$ uniformly in x.

We now use the canonical map to construct a generator of $\pi_n(R_n, R_{n-1})$. We shall take as fixed reference points for this group the north pole of S^{n-1} and the identity matrix $I \in R_{n-1}$. Since ([2], Theorem 2) $\pi_n(R_n, R_{n-1})$ is isomorphic to $\pi_n(S^n)$ under the map $\pi(R_n) = S^n$, a generator of the former group is represented by any relative n-cell in R_n modulo R_{n-1} which projects into S^n with degree ± 1. To define such a relative n-cell, we observe that C_n maps V_1^n into R_n and S^{n-1} into the coset $\tilde{R}_{n-1}$ opposite to R_{n-1} and projects into S^n with degree 1. Hence the map $D_n(x) = C_n(x) \cdot \begin{Vmatrix} I_{n-1} & 0 \\ 0 & -I_2 \end{Vmatrix}$ $(x \in V_1^n)$ defines a relative cell in R_n modulo R_{n-1}, and it is easily verified that $D_n(x^0) = I$, while $d_n(x) = \pi D_n(x) = \tilde{g}_n(x)$ has degree $(-1)^{n+1}$. Hence D_n represents the required generator.

Since $\pi_{n-1}(R_n, R_{n-1}) = \pi_{n-1}(S^n) = 0$, it follows from the results of Section 3 that $\pi_{n-1}(R_n) = \pi_{n-1}(R_{n-1})/\pi_{n-1,o}(R_{n-1})$. Thus $\pi_{n-1}(R_n)$ is a factor group of $\pi_{n-1}(R_{n-1})$, the kernel of the homomorphism being the group $\pi_{n-1,o}(R_{n-1})$. But $\pi_{n-1,o}(R_{n-1}) = \omega[\pi_n(R_n, R_{n-1})]$; hence a generator of $\pi_{n-1,o}(R_{n-1})$ is given by the map ωD_n, which is easily shown to be the canonical map C_{n-1}. Hence

THEOREM 5. *The kernel of the homomorphism $\pi_{n-1}(R_{n-1}) \to \pi_{n-1}(R_n)$ is the subgroup of the former group generated by the canonical map.*

7. On the possibility of sectioning the cosets of R_{n-1} in R_n

In this section the following question is considered: Is there an n-sphere in R_n which projects into S^n with degree 1? It is shown that this question can be answered in the negative for certain values of n. For other dimensions the question remains open.

The first step toward the solution of this problem appears in

THEOREM 6. *The following conditions are equivalent:*

1) *there is a map $F(S^n) \subset R_n$ such that πF has degree one;*
2) *R_n can be represented as a product space $R_{n-1} \times S^n$;*
3) *the homomorphism $\pi_{n-1}(R_{n-1}) \to \pi_{n-1}(R_n)$ is an isomorphism;*
4) *the canonical map of S^{n-1} into R_{n-1} is homotopic to a point in R_{n-1}.*

The first condition implies the second. For let $F(S^n) \subset R_n$ be such that πF has degree one. Since πF is homotopic to the identity, it follows from the covering homotopy theorem ([2], Theorem 1) that F is homotopic to a map $F'(S^n) \subset R_n$ such that $\pi F'(x) = x$. Then by Theorem 2, $R_n = R_{n-1} \times S^n$.

That the second condition implies the third we have observed in the proof of the Corollary to Theorem 3; that the third implies the fourth follows from Theorem 5.

The fourth implies the first; for during the homotopy of C_{n-1} to a point a relative n-cell is swept out in R_{n-1} whose boundary is the $(n-1)$-sphere defined by the canonical map. But the map D_n defines a relative n-cell in R_n with the same boundary as the first one. Joining these two cells by identifying corre-

sponding points on the boundaries, we obtain a sphere in R_n whose projection has the same degree $(-1)^{n+1}$ as that of d_n. The required map is then easily constructed.

Theorem 6 enables us to answer immediately the question posed above for the case when n is even. For suppose that n is even and suppose that such a map exists. Then by the fourth condition in Theorem 6, the canonical map C_{n-1} is homotopic to a point in R_{n-1}. Hence $g_{n-1} = \pi C_{n-1}$ is inessential. But we have already shown that g_{n-1} has degree two. This contradiction completes the proof of

THEOREM 7. *If n is even, there is no map $F(S^n) \subset R_n$ such that πF has degree one.*

We now turn to the much more difficult case where n is odd. A partial answer to the question is obtained in

THEOREM 8. *If $n > 1$ and $n \equiv 1 \pmod 4$, there is no map $F(S^n) \subset R_n$ such that πF has degree one.*

The proof may be outlined as follows:

1) Since for n odd, g_{n-1} is homotopic to a point, it follows that C_{n-1} is homotopic to a map $G_{n-1}(S^{n-1}) \subset R_{n-2}$. Such a homotopy is exhibited, and it is proved that $k_{n-1} = \pi G_{n-1}$ is essential if $n \equiv 1 \pmod 4$, and inessential if $n \equiv 3 \pmod 4$.

2) A generator $P_n(V_1^n) \subset R_{n-1}$ of the group $\pi_n(R_{n-1}, R_{n-2})$ is constructed for all $n \geqq 5$, and the projection p_{n-1}^* of the map $\omega P_n(S^{n-1}) \subset R_{n-2}$ is shown to be inessential.

When these steps have been established, the proof of the theorem may be completed as follows: Suppose that $n > 1$ and $n \equiv 1 \pmod 4$ and suppose that the theorem is not true. Then C_{n-1}, and consequently G_{n-1}, is homotopic to a point in R_{n-1}. Since G_{n-1} is not homotopic to a point in R_{n-2}, the deformation of G_{n-1} defines a relative n-cell in R_{n-1} modulo R_{n-2}, and hence an essential element of $\pi_n(R_{n-1}, R_{n-2})$. This group has been shown by Freudenthal [3] to be the cyclic group of period 2 if $n \geqq 4$. Hence ωP_n and G_{n-1} are homotopic in R_{n-2}. But p_{n-1}^* and k_{n-1} are not homotopic, a contradiction.

LEMMA. *If n is odd, the canonical map C_{n-1} is homotopic in R_{n-1} to a map $G_{n-1}(S^{n-1}) \subset R_{n-2}$, whose projection into S^{n-2} is essential if $n \equiv 1 \pmod 4$ and inessential if $n \equiv 3 \pmod 4$.*

Let E^n denote the closed n-cell bounded by the unit sphere S^{n-1} in Euclidean n-space. H^n denotes the upper half $x_n \geqq 0$ of E^n. Let points $y \in H^n$ be represented by coördinates (x, r) where $x \in V_1^{n-1}$ is the central projection of the point y on V_1^{n-1}, and r is the distance of the point y from the origin. Let

$$H(r) = \begin{Vmatrix} c & s & 0 & 0 & \cdots & 0 \\ -s & c & 0 & 0 & \cdots & 0 \\ 0 & 0 & c & s & \cdots & 0 \\ 0 & 0 & -s & c & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{Vmatrix},$$

where $c = 1 - 2r^2$, $s = 2r(1 - r^2)^{\frac{1}{2}}$ $(0 \leqq r \leqq 1)$. Let

$$G(y) = G(x, r) = C_{n-1}(x) \cdot H(r) \cdot C_{n-1}(x)^{-1} \cdot \begin{Vmatrix} -I_{n-1} & 0 \\ 0 & 1 \end{Vmatrix} \tag{6}$$

$$(x \in V_1^n; 0 \leqq r \leqq 1)$$

map H^n into R_{n-1}. Evidently $G(x) = C_{n-1}[g_{n-1}(x)]$ for $x \in V_1^{n-1}$, while G map, the equatorial plane into R_{n-2} and S^{n-2} into a single point. Deform V_1^{n-1} over H^n into E^{n-1} by letting each point move along the perpendicular joining it to the plane $x_n = 0$. The image of G under this homotopy gives a homotopy of $C_{n-1}[g_{n-1}(x)]$ to a map $K(V_1^{n-1}) \subset R_{n-2}$. Since under the homotopy S^{n-2} remains fixed, K maps S^{n-2} into a single point. Evidently $K(x)$ can be represented in the form $K(x) = G_{n-1}[g_{n-1}(x)]$, where $G_{n-1}(S^{n-1}) \subset R_{n-2}$ is defined and continuous over all of S^{n-1}, and G_{n-1} is homotopic in R_{n-1} to C_{n-1}.

By multiplying out the matrices in (6) and computing the homotopy, we find that the map $k_{n-1} = \pi G_{n-1}$ of S^{n-1} into S^{n-2} is represented by the equations

$$\begin{aligned}
y_{2i-1} &= \frac{2(x_{2i-1}x_{n-2} + x_{2i}x_{n-1})}{(1 - x_n^2)^{\frac{1}{2}}}, \\
y_{2i} &= \frac{2(x_{2i}x_{n-2} - x_{2i-1}x_{n-1})}{(1 - x_n^2)^{\frac{1}{2}}} \qquad \left(i = 1, \cdots, \frac{n-3}{2}\right), \\
y_{n-2} &= \frac{2(x_{n-2}^2 + x_{n-1}^2)}{(1 - x_n^2)^{\frac{1}{2}}} - (1 - x_n^2)^{\frac{1}{2}}, \\
y_{n-1} &= x_n \qquad (1 - x_n^2 \neq 0); \\
y_i &= 0 \ (i = 1, \cdots, n-2), \qquad y_{n-1} = x_n \ (1 - x_n^2 = 0).
\end{aligned} \tag{7}$$

It follows easily that k_{n-1} is defined and continuous over all of S^{n-1} and maps S^{n-1} on S^{n-2}. In order to investigate this map, let us consider the map $m_{n-1}(S^{n-2}) = S^{n-3}$ obtained by setting $x_n = y_{n-1} = 0$ in (7). This map can be studied more easily in complex coördinates. Let $z_j = x_{2j-1} + ix_{2j}$, $w_j = y_{2j-1} + iy_{2j}$ $(j = 1, \cdots, (n-1)/2 = k)$, where $w_k = y_{n-2}$ is real. Then S^{n-2} and S^{n-3} are given by the equations $\sum z_j\bar{z}_j = 1$, $\sum w_j\bar{w}_j = 1$, respectively. In terms of these coördinates the equations representing the map m_{n-1} take the form

$$\begin{aligned}
w_j &= 2z_j\bar{z}_k \qquad (j = 1, \cdots, k-1), \\
w_k &= 2z_k\bar{z}_k - 1.
\end{aligned} \tag{8}$$

If the complex coördinates occurring in (8) are formally replaced by real ones, the map g_{k-1} is obtained. In equations (3) and (5) we have constructed homotopies of g_n for n even and odd, respectively. The functions defining these homotopies can be extended so as to be defined for complex coördinates, as follows: if k is odd, i.e., if $n \equiv 3 \pmod 4$, let

$$w_{2j-1} = 2\{(1 - t)z_{2j-1}\bar{z}_k + [t(1 - t)]^{\frac{1}{2}}\bar{z}_{2j}\},$$

$$(9)\qquad w_{2j} = 2\{(1-t)z_{2j}\bar{z}_k - [t(1-t)]^{\frac{1}{2}}\bar{z}_{2j-1}\} \qquad \left(j = 1, \cdots, \frac{k-1}{2}\right),$$

$$w_k = 1 - 2(1-t)(1 - z_k\bar{z}_k).$$

Evidently the homotopy (9) deforms m_{n-1} over S^{n-3} to a point. Hence m_{n-1}, and consequently also k_{n-1}, is homotopic to a point. On the other hand, if k is even, i.e., if $n \equiv 1 \pmod 4$, let us consider the map $m'_{n-1}(S^{n-2}) \subset S^{n-3}$ defined by the equations

$$w_j = z_j \qquad (j = 1, \cdots, k-2),$$

$$(10)\qquad w_{k-1} = \frac{2z_{k-1}\bar{z}_k}{(z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)^{\frac{1}{2}}},$$

$$w_k = \frac{2z_k\bar{z}_k}{(z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)^{\frac{1}{2}}} - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)^{\frac{1}{2}} \qquad (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k \neq 0);$$

$$w_{k-1} = w_k = 0 \quad \text{for} \quad z_{k-1} = z_k = 0.$$

Setting $w_j = z_j = 0$ $(j = 1, \cdots, k \equiv 2)$ in (10) and writing the result in real coördinates, we obtain a map of S^3 on S^2 of Hopf invariant ± 1 (cf. Hopf [4], §5). It follows from the results of Freudenthal [3] that m'_{n-1} is essential.

We shall now define a deformation of m_{n-1} to m'_{n-1} (for $n \equiv 1 \pmod 4$) by extending the functions in (5) so as to be defined for complex coördinates, as follows: let

$$w_{2j-1} = tz_{2j-1} + [t(1-t)]^{\frac{1}{2}}\bar{z}_{2j} + 2\frac{(1-t)z_{2j-1}\bar{z}_k - [t(1-t)]^{\frac{1}{2}}z_k\bar{z}_{2j}}{\{1 - t[1 - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)]\}^{\frac{1}{2}}},$$

$$w_{2j} = tz_{2j} - [t(1-t)]^{\frac{1}{2}}\bar{z}_{2j-1} + 2\frac{(1-t)z_{2j}\bar{z}_k + [t(1-t)]^{\frac{1}{2}}z_k\bar{z}_{2j-1}}{\{1 - t[1 - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)]\}^{\frac{1}{2}}}$$

$$(11)\qquad \left(j = 1, \cdots, \frac{k-2}{2}\right),$$

$$w_{k-1} = \frac{2z_{k-1}\bar{z}_k}{\{1 - t[1 - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)]\}^{\frac{1}{2}}},$$

$$w_k = \frac{2z_k\bar{z}_k}{\{1 - t[1 - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)]\}^{\frac{1}{2}}} - \{1 - t[1 - (z_{k-1}\bar{z}_{k-1} + z_k\bar{z}_k)]\}^{\frac{1}{2}}.$$

Although this map is not defined everywhere for $t = 1$, it is readily verified that as $t \to 1$, the functions in (11) converge to the corresponding ones in (10) uniformly in z. Thus m_{n-1} and m'_{n-1} are homotopic, so that m_{n-1}, and consequently also k_{n-1}, is essential. This completes the proof of the Lemma.

As indicated above, the next step in the proof is to construct a generator of the group $\pi_n(R_{n-1}, R_{n-2})$. This group is isomorphic with $\pi_n(S^{n-1})$ under the projection π ([2], Theorem 2). If $n \geq 4$ we may take as generator of the latter group the map $p_n(S^n) = S^{n-1}$ defined by

$$p_{n,i}(x) = x_i \qquad (i = 1, \cdots, n-4),$$

$$
\begin{aligned}
p_{n,n-3}(x) &= \frac{2(x_{n-3}x_{n-1} + x_{n-2}x_n)}{(x_n^2 + x_{n-1}^2 + x_{n-2}^2 + x_{n-3}^2)^{\frac{1}{2}}}, \\
p_{n,n-2}(x) &= \frac{2(x_{n-2}x_{n-1} - x_{n-3}x_n)}{(x_n^2 + x_{n-1}^2 + x_{n-2}^2 + x_{n-3}^2)^{\frac{1}{2}}}, \\
p_{n,n-1}(x) &= \frac{x_n^2 + x_{n-1}^2 - x_{n-2}^2 - x_{n-3}^2}{(x_n^2 + x_{n-1}^2 + x_{n-2}^2 + x_{n-3}^2)^{\frac{1}{2}}}, \\
p_{n,n}(x) &= x_{n+1} \qquad (x_n^2 + x_{n-1}^2 + x_{n-2}^2 + x_{n-3}^2 \neq 0), \\
p_{n,n-3}(x) &= p_{n,n-2}(x) = p_{n,n-1}(x) = 0 \qquad (x_n = x_{n-1} = x_{n-2} = x_{n-3} = 0).
\end{aligned} \tag{12}
$$

Evidently p_n maps V_1^n into V_1^{n-1} and S^{n-1} into S^{n-2}. The latter map is essential; in fact, it represents a generator of $\pi_{n-1}(S^{n-2})$. Let $P_n(x) = D_{n-1}[p_n(x)]$ $(x \in V_1^n)$. Then P_n represents the desired generator of $\pi_n(R_{n-1}, R_{n-2})$; for $\pi P_n(x) = d_{n-1}[p_n(x)]$, and the latter map represents a generator of $\pi_n(S^{n-1}, x^0) = \pi_n(S^{n-1})$.

Let $P_n^* = \omega P_n$, and $p_{n-1}^* = \pi P_n^*$ its projection into S^{n-2}. Since p_n maps S^{n-1} into S^{n-2}, and since $\omega D_{n-1} = C_{n-2}$, we have $P_n^*(x) = C_{n-2}[p_n(x)]$, and hence $p_{n-1}^*(x) = g_{n-2}[p_n(x)]$ $(x \in S^{n-1})$. Let α denote the element of $\pi_{n-2}(S^{n-2})$ determined by the identity map, β the element of $\pi_{n-1}(S^{n-2})$ defined by the map $p_n(S^{n-1}) = S^{n-2}$. Since $\{g_{n-2}\} = 0$ or 2α according as n is even or odd, we have $\{p_{n-1}^*\} = 0$ or 4β [2] respectively. But Freudenthal [3] has proved that $2\beta = 0$ if $n \geqq 5$; hence for all $n \geqq 5$ we have $\{p_{n-1}^*\} = 0$, so that p_{n-1}^* is inessential. This completes the proof of the theorem.

8. Computation of the homotopy groups

We are now in a position to establish the results exhibited in Section 2. The generators which we shall compute here will differ slightly from those already exhibited; however, they may be easily seen to be homotopic.

We have already observed that $\pi_n(R_{n+1})$ is a homomorphic image of $\pi_n(R_n)$. In a similar manner we can prove (cf. [2], Theorem 5) that $\pi_n(R_{n+k})$ $(k = 1, 2, \cdots)$ is isomorphic with $\pi_n(R_{n+1})$. Another useful result is the following:

THEOREM 9. *If n is even, $\pi_n(R_n)$ is a factor group of $\pi_n(R_{n-1})$; the kernel of the homomorphism is the group $\pi_{n,0}(R_{n-1})$.*

Since $\{D_n\}$ is a generator of $\pi_n(R_n, R_{n-1})$ and since $\omega\{D_n\} = \{C_{n-1}\} \neq 0$ for n even, it follows that $\pi_{n,0}(R_n, R_{n-1}) = 0$. Hence $\pi_n(R_n) = \pi_n(R_{n-1})/\pi_{n,0}(R_{n-1})$.

In order to compute $\pi_1(R_n)$, we first observe that, since R_1 is homeomorphic to S^1, its fundamental group is the free cyclic group generated by any map of S^1 on R_1 of degree 1. Such a generator α_1 is given by the map

$$
B_1(x) = \begin{Vmatrix} x_2 & x_1 \\ -x_1 & x_2 \end{Vmatrix}.
$$

Since R_2 is homeomorphic with projective 3-space P^3, it follows that $\pi_1(R_2)$

[2] This follows easily from Theorem II b′ of [4].

is the cyclic group of order two. But $\pi_1(R_2)$ is a factor group of $\pi_1(R_1)$; hence for a generator of $\pi_1(R_2)$ we may take the generator α_1 for $\pi_1(R_1)$ subject to the condition $2\alpha_1 = 0$ in R_2. The same result holds for $\pi_1(R_n)$.

Since all the higher homotopy groups of S^1 vanish, the same holds for R_1. In particular, $\pi_2(R_1) = 0$. Then, by Theorem 9, $\pi_2(R_n) = 0$ for all n.

The higher homotopy groups of R_2 are isomorphic to those of its covering space S^3. In particular, $\pi_3(R_2)$ is the free cyclic group, and a generator α_3 is represented by the covering map $H(S^3) = R_2$ given by the equation

$$H(x) = \begin{Vmatrix} x_4^2 - x_3^2 - x_2^2 + x_1^2 & 2(x_1x_2 - x_3x_4) & 2(x_1x_3 + x_2x_4) \\ 2(x_1x_3 + x_2x_4) & x_4^2 - x_3^2 + x_2^2 - x_1^2 & 2(x_2x_3 - x_1x_4) \\ 2(x_1x_3 - x_2x_4) & 2(x_1x_4 + x_2x_3) & x_4^2 + x_3^2 - x_2^2 - x_1^2 \end{Vmatrix}.$$

We observe that $\pi H(x)$ maps S^3 on S^2 with Hopf invariant ± 1.

Since R_3 is the product space of R_2 and the quaternion subgroup Q^3, the homotopy groups of R_3 are the direct sums of the groups of the same dimension of R_2 and $Q^3 = S^3$. In particular, $\pi_3(R_3) = \pi_3(R_2) + \pi_3(S^3)$ is a free group with two generators. For one of these generators we may take the generator α_3 of $\pi_3(R_2)$; the second generator β_3 is determined by the quaternion matrix

$$B_3(x) = \begin{Vmatrix} x_4 & x_3 & -x_2 & x_1 \\ -x_3 & x_4 & x_1 & x_2 \\ -x_2 & -x_1 & x_4 & x_3 \\ -x_1 & -x_2 & -x_3 & x_4 \end{Vmatrix} \qquad (x \in S^3).$$

To compute $\pi_3(R_4) = \pi_3(R_3)/\pi_{3,o}(R_3)$, we observe that

$$C_3(x) = B_3[g_3(x)] \cdot \begin{Vmatrix} H(x) & 0 \\ 0 & 1 \end{Vmatrix}$$

so that $\{C_3\} = 2\beta_3 + \alpha_3 = 0$ in R_4. Hence $\pi_3(R_4)$ is a free cyclic group with the one generator β_3, and the same is true of $\pi_3(R_n)$ $(n \geqq 4)$.

We now consider the groups π_4. Freudenthal [3] and Pontrjagin [5] have proved that $\pi_4(S^3)$ is the cyclic group of order two, a generator being determined by the map $p_4(S^4) = S^3$. Hence a generator α_4 of $\pi_4(R_2)$ is determined by the map Hp_4. The group $\pi_4(R_3)$ is the direct sum of two cyclic groups of order two; for generators we may take α_4 and $\beta_4 = \{B_3p_4\}$.

We have observed in the proof of Theorem 8 that $\pi_{4,o}(R_3)$ is generated by $\omega\{P_5\} = \{P_5^*\}$. But

$$P_5^*(x) = B_3[p_4^*(x)] \cdot \begin{Vmatrix} H[p_4(x)] & 0 \\ 0 & 1 \end{Vmatrix}$$

and $\{p_4^*\} = 0$, so that $\{P_5^*\} = \alpha_4 = 0$ in R_4. Hence $\pi_4(R_4)$ is the cyclic group of order two generated by β_4.

We have proved in Theorem 8 that $\{C_4\} \neq 0$ in R_4. Hence $\pi_{4,o}(R_4) = \pi_4(R_4)$, so that $\pi_4(R_5) = \pi_4(R_n) = 0$ $(n \geqq 5)$.

We conclude by computing the five-dimensional homotopy groups of R_n. Pontrjagin [5] has proved that $\pi_5(S^3) = 0$; hence $\pi_5(R_2) = \pi_5(R_3) = 0$. Hence $\pi_5(R_4) = \pi_{5,o}(R_4, R_3) = 0$ since $\pi_{4,o}(R_3) \neq 0$. This in turn implies that $\pi_5(R_5) = \pi_{5,o}(R_5, R_4)$. But $\pi_5(R_5)$ contains the essential element $\alpha_5 = \{C_5\}$, and since πC_5 has degree two, it follows from Theorem 8 that $\pi_5(R_5)$ is the free cyclic group generated by α_5. Since $\alpha_5 = 0$ in R_6, we have finally that $\pi_5(R_6) = \pi_5(R_n) = 0$ $(n \geqq 6)$.

9. Continuous vector fields over spheres

The above results will now be applied to the study of continuous vector fields over S^n. Geometrically, a continuous vector field may be thought of as a set of functions $V^i(x)$ $(i = 1, \cdots, n+1; x \in S^n)$ defining a unit vector tangent to S^n at the point x, the functions $V^i(x)$ being continuous over all of S^n. A set of p such fields are said to be *independent* if the vectors of the fields at each point of S^n are independent vectors.

Let P_k denote the point of S^n whose coördinates are δ_{ik} $(i, k = 1, \cdots, n+1)$. R_{n-p} denotes the subgroup of R_n consisting of all rotations leaving P_k fixed $(k = n-p+2, \cdots, n+1)$. The coset space R_n/R_{n-p} may be thought of as the space of all sets of p mutually orthogonal points of S^n; for two elements of R_n are in the same coset of R_{n-p} if and only if their last p columns are identical. These columns define the orthogonal p-tuple associated with the given coset. Conversely, given a set of orthogonal points $Q_1, Q_2, \cdots, Q_p$, the required coset is the set of all rotations carrying P_{n-p+2} into $Q_1, \cdots$, and P_{n+1} into Q_p.

Since $R_{n-p} \subset R_{n-1}$, each coset of R_{n-p} lies in some coset of R_{n-1}. Thus a natural mapping $R_n/R_{n-p} \to R_n/R_{n-1} = S^n$ is defined; each coset of R_{n-p} is mapped into the coset of R_{n-1} containing it. We refer to this map as the *projection*; it is easily verified that it is a fibre map. If we regard an element of R_n/R_{n-p} as being determined by a set of p mutually orthogonal points, the projection is the p^{th} point of the set.

Since the usual process of orthogonalizing a set of independent vectors can be carried out here, there is no loss of generality in assuming that the vectors of the fields we are dealing with are orthogonal at each point and of unit length.

THEOREM 10. *Every set of p orthogonal vector fields over S^n defines a mapping of S^n into R_n/R_{n-p-1} whose projection into S^n is the identity; conversely, any such map defines a set of p orthogonal vector fields.*

By translating the vectors at any point x to the center of S^n and adjoining the point x, we obtain an orthogonal $(p+1)$-tuple whose projection is x. This process is continuous and gives the required map. Conversely, given such a map, we define the p orthogonal vectors at x as follows: the given map associates with x an orthogonal $(p+1)$-tuple, the $(p+1)^{\text{st}}$ point of which is x. Transla-

tion of radius vectors drawn to the first p points to the point x gives the set of vectors at x.

COROLLARY. *There exists a set of p independent vector fields over S^n if and only if the canonical map of S^{n-1} into R_{n-1} is contractible in R_{n-1} into R_{n-p-1}.*

It follows from Theorem 7 that there is no vector field over S^n if n is even. However, if n is odd, it follows from the Lemma used in the proof of Theorem 8 that there is at least one. Over S^3 and S^7 are three and seven orthogonal fields, respectively. These are defined by the mappings $S^3 \to R_3 \to S^3$ and $S^7 \to R_7 \to S^7$ of degree one given by the matrices B and B obtained in Section 5.

THEOREM 11. *If $n \equiv 3 \pmod 4$ there at least three orthogonal vector fields over S^n; if $n \equiv 7 \pmod 8$ there are at least seven.*

We shall prove the theorem for $n = 4m + 3$; the proof for $n = 8m + 7$ is entirely analogous. Let $V^i(x)$ $(i = 1, 2, 3;\ x \in S^n)$ define a set of orthogonal vector fields over S^n; the components of the vector $V^i(x)$ will be denoted by $V^i_j(x)$ $(j = 1, 2, 3, 4)$. We extend $V^i(x)$ to be defined over all of E^4 as follows: if $y \in E^4$, let x be the central projection of y on S^3, r the distance of y from the origin. Then $V^i(y) = rV^i(x)$ defines a set of three orthogonal vectors at each point of E^4, and these vectors vanish only at the origin.

If z is a point of E^{4m+4} with coördinates $(z_1, \cdots, z_{4m+4})$, let x^j $(j = 1, \cdots, m + 1)$ denote the point of E^4 with coördinates $(z_{4j-3}, z_{4j-2}, z_{4j-1}, z_{4j})$. Then we define three vector fields $W^i(z)$ over E^{4m+4} as follows:

$$W^i_{4j+k}(z) = V^i_k(x^{j+1}) \qquad (i = 1, 2, 3;\ j = 0, \cdots, m;\ k = 1, 2, 3, 4).$$

Evidently the vectors $W^i(z)$ are mutually orthogonal at each point z, and if $z \in S^{4m+3}$ they are of unit length. Thus the theorem is proved.

The problem of determining the maximum number of independent fields which can exist over S^n has not yet been solved for general odd n. For $n \equiv 1 \pmod 4$ the solution appears in

THEOREM 12. *If $n \equiv 1 \pmod 4$ any two vector fields over S^n are somewhere dependent.*

The theorem is evidently true for $n = 1$. Suppose $n > 1$ and suppose that the theorem were not true. Then by the Corollary to Theorem 10, the canonical map C_{n-1} would be contractible in R_{n-1} into R_{n-3}. Hence the map $G_{n-1}(S^{n-1}) \subset R_{n-2}$ would be homotopic in R_{n-1} to a map $F(S^{n-1}) \subset R_{n-3}$. Hence $\{G_{n-1}\} - \{F\} \in \pi_{n-1,o}(R_{n-2})$. But $\pi[\{G_{n-1}\} - \{F\}] = \pi\{G_{n-1}\} - \pi\{F\} = \pi\{G_{n-1}\} \neq 0$. In the proof of Theorem 8 we have shown that this is impossible.

COROLLARY. *If $n > 1$ and $n \equiv 1 \pmod 4$ the tangent sphere-bundle of S^n is not simple* ([7], p. 788).

UNIVERSITY OF CHICAGO

BIBLIOGRAPHY

1. W. HUREWICZ, *Beiträge zur Topologie der Deformationen*, Proc. Kon. Akad. van Wetenschappen te Amsterdam, 38 (1935), pp. 112–119, 521–528; 39 (1936), pp. 117–125, 215–224.

We have proved in Theorem 8 that $\{C_4\} \neq 0$ in R_4. Hence $\pi_{4,o}(R_4) = \pi_4(R_4)$, so that $\pi_4(R_5) = \pi_4(R_n) = 0$ $(n \geqq 5)$.

We conclude by computing the five-dimensional homotopy groups of R_n. Pontrjagin [5] has proved that $\pi_5(S^3) = 0$; hence $\pi_5(R_2) = \pi_5(R_3) = 0$. Hence $\pi_5(R_4) = \pi_{5,o}(R_4, R_3) = 0$ since $\pi_{4,o}(R_3) \neq 0$. This in turn implies that $\pi_5(R_5) = \pi_{5,o}(R_5, R_4)$. But $\pi_5(R_5)$ contains the essential element $\alpha_5 = \{C_5\}$, and since πC_5 has degree two, it follows from Theorem 8 that $\pi_5(R_5)$ is the free cyclic group generated by α_5. Since $\alpha_5 = 0$ in R_6, we have finally that $\pi_5(R_6) = \pi_5(R_n) = 0$ $(n \geqq 6)$.

9. Continuous vector fields over spheres

The above results will now be applied to the study of continuous vector fields over S^n. Geometrically, a continuous vector field may be thought of as a set of functions $V^i(x)$ $(i = 1, \cdots, n+1; x \in S^n)$ defining a unit vector tangent to S^n at the point x, the functions $V^i(x)$ being continuous over all of S^n. A set of p such fields are said to be *independent* if the vectors of the fields at each point of S^n are independent vectors.

Let P_k denote the point of S^n whose coördinates are δ_{ik} $(i, k = 1, \cdots, n+1)$. R_{n-p} denotes the subgroup of R_n consisting of all rotations leaving P_k fixed $(k = n-p+2, \cdots, n+1)$. The coset space R_n/R_{n-p} may be thought of as the space of all sets of p mutually orthogonal points of S^n; for two elements of R_n are in the same coset of R_{n-p} if and only if their last p columns are identical. These columns define the orthogonal p-tuple associated with the given coset. Conversely, given a set of orthogonal points $Q_1, Q_2, \cdots, Q_p$, the required coset is the set of all rotations carrying P_{n-p+2} into $Q_1, \cdots$, and P_{n+1} into Q_p.

Since $R_{n-p} \subset R_{n-1}$, each coset of R_{n-p} lies in some coset of R_{n-1}. Thus a natural mapping $R_n/R_{n-p} \to R_n/R_{n-1} = S^n$ is defined; each coset of R_{n-p} is mapped into the coset of R_{n-1} containing it. We refer to this map as the *projection*; it is easily verified that it is a fibre map. If we regard an element of R_n/R_{n-p} as being determined by a set of p mutually orthogonal points, the projection is the p^{th} point of the set.

Since the usual process of orthogonalizing a set of independent vectors can be carried out here, there is no loss of generality in assuming that the vectors of the fields we are dealing with are orthogonal at each point and of unit length.

THEOREM 10. *Every set of p orthogonal vector fields over S^n defines a mapping of S^n into R_n/R_{n-p-1} whose projection into S^n is the identity; conversely, any such map defines a set of p orthogonal vector fields.*

By translating the vectors at any point x to the center of S^n and adjoining the point x, we obtain an orthogonal $(p+1)$-tuple whose projection is x. This process is continuous and gives the required map. Conversely, given such a map, we define the p orthogonal vectors at x as follows: the given map associates with x an orthogonal $(p+1)$-tuple, the $(p+1)^{\text{st}}$ point of which is x. Transla-

tion of radius vectors drawn to the first p points to the point x gives the set of vectors at x.

COROLLARY. *There exists a set of p independent vector fields over S^n if and only if the canonical map of S^{n-1} into R_{n-1} is contractible in R_{n-1} into R_{n-p-1}.*

It follows from Theorem 7 that there is no vector field over S^n if n is even. However, if n is odd, it follows from the Lemma used in the proof of Theorem 8 that there is at least one. Over S^3 and S^7 are three and seven orthogonal fields, respectively. These are defined by the mappings $S^3 \to R_3 \to S^3$ and $S^7 \to R_7 \to S^7$ of degree one given by the matrices B and B obtained in Section 5.

THEOREM 11. *If $n \equiv 3 \pmod 4$ there at least three orthogonal vector fields over S^n; if $n \equiv 7 \pmod 8$ there are at least seven.*

We shall prove the theorem for $n = 4m + 3$; the proof for $n = 8m + 7$ is entirely analogous. Let $V^i(x)$ $(i = 1, 2, 3;\ x \in S^n)$ define a set of orthogonal vector fields over S^n; the components of the vector $V^i(x)$ will be denoted by $V^i_j(x)$ $(j = 1, 2, 3, 4)$. We extend $V^i(x)$ to be defined over all of E^4 as follows: if $y \in E^4$, let x be the central projection of y on S^3, r the distance of y from the origin. Then $V^i(y) = rV^i(x)$ defines a set of three orthogonal vectors at each point of E^4, and these vectors vanish only at the origin.

If z is a point of E^{4m+4} with coördinates $(z_1, \cdots, z_{4m+4})$, let x^j $(j = 1, \cdots, m + 1)$ denote the point of E^4 with coördinates $(z_{4j-3}, z_{4j-2}, z_{4j-1}, z_{4j})$. Then we define three vector fields $W^i(z)$ over E^{4m+4} as follows:

$$W^i_{4j+k}(z) = V^i_k(x^{j+1}) \qquad (i = 1, 2, 3;\ j = 0, \cdots, m;\ k = 1, 2, 3, 4).$$

Evidently the vectors $W^i(z)$ are mutually orthogonal at each point z, and if $z \in S^{4m+3}$ they are of unit length. Thus the theorem is proved.

The problem of determining the maximum number of independent fields which can exist over S^n has not yet been solved for general odd n. For $n \equiv 1 \pmod 4$ the solution appears in

THEOREM 12. *If $n \equiv 1 \pmod 4$ any two vector fields over S^n are somewhere dependent.*

The theorem is evidently true for $n = 1$. Suppose $n > 1$ and suppose that the theorem were not true. Then by the Corollary to Theorem 10, the canonical map C_{n-1} would be contractible in R_{n-1} into R_{n-3}. Hence the map $G_{n-1}(S^{n-1}) \subset R_{n-2}$ would be homotopic in R_{n-1} to a map $F(S^{n-1}) \subset R_{n-3}$. Hence $\{G_{n-1}\} - \{F\} \in \pi_{n-1,o}(R_{n-2})$. But $\pi[\{G_{n-1}\} - \{F\}] = \pi\{G_{n-1}\} - \pi\{F\} = \pi\{G_{n-1}\} \neq 0$. In the proof of Theorem 8 we have shown that this is impossible.

COROLLARY. *If $n > 1$ and $n \equiv 1 \pmod 4$ the tangent sphere-bundle of S^n is not simple* ([7], p. 788).

UNIVERSITY OF CHICAGO

BIBLIOGRAPHY

1. W. HUREWICZ, *Beiträge zur Topologie der Deformationen*, Proc. Kon. Akad. van Wetenschappen te Amsterdam, 38 (1935), pp. 112–119, 521–528; 39 (1936), pp. 117–125, 215–224.

2. W. Hurewicz and N. E. Steenrod, *Homotopy relations in fibre spaces*, Proc. Nat. Acad. of Sci., 27 (1941), pp. 61–64.
3. H. Freudenthal, *Über die Klassen der Sphärenabbildungen*, Comp. Math., 5 (1937), pp. 299–314.
4. H. Hopf, *Über die Abbildungen der dreidimensionalen Sphäre auf die Kugelfläche*, Math. Ann., 104 (1931), pp. 637–665.
5. L. Pontrjagin, *A classification of continuous transformations of a complex into a sphere*, C. R., Acad. des Sc. de l'URSS, 19 (1938), pp. 361–363.
6. H. Whitney, *On the theory of sphere-bundles*, Proc. Nat. Acad. of Sci., 26 (1940), pp. 148–153.
7. H. Whitney, *Topological properties of differentiable manifolds*, Bull. Am. Math. Soc., 43 (1937), pp. 785–805.
8. E. Cartan, *La topologie des groupes de Lie*, Actualités Scientifiques et Industrielles, Exposés de geométrie, vol. VIII. Paris, 1936.

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ON MATRIX ALGEBRAS OVER AN ALGEBRAICALLY CLOSED FIELD*†

BY WINSTON M. SCOTT

(Received June 2, 1941)

Introduction

Recently a number of writers have discussed interesting developments in the theory of not completely reducible matrix sets and non-semisimple algebras.[1] Here we have made use of some of these concepts and methods to study matrix algebras over an algebraically closed field. We shall discuss in some detail in this introduction the definitions that are used and the theorems that are developed.

Let $\mathfrak{A}$ denote a matrix algebra, with unit element, over an algebraically closed field K. $\mathfrak{A}$ will be taken in reduced form, by which we shall mean that $\mathfrak{A}$ is exhibited with only zeros above the main diagonal, with irreducible constituents of $\mathfrak{A}$ in the main diagonal, and that $\mathfrak{A}$ is expressible as a direct sum of its radical and a semisimple subalgebra which latter has non-zero components only in the irreducible constituents of $\mathfrak{A}$:[2]

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{C}_{11} & & & \\ \mathfrak{C}_{21} & \mathfrak{C}_{22} & & \\ \vdots & & \ddots & \\ \mathfrak{C}_{t1} & \mathfrak{C}_{t2} & \cdots & \mathfrak{C}_{tt} \end{pmatrix}, \tag{1}$$

the $\mathfrak{C}_{ii}$ denoting the irreducible constituents; further

$$\mathfrak{A} = \mathfrak{N} + \mathfrak{A}^*,$$

---

* The following constitutes a portion of a dissertation written under the direction of Dr. C. J. Nesbitt and accepted by the University of Michigan in January, 1941. The aid and encouragement given me by Dr. Nesbitt has been invaluable in the preparation of this work.

† Presented to the Society, May 2, 1941.

[1] See, for instance, the forthcoming paper by R. Brauer, *On Sets of Matrices with Coefficients in a Division Ring*; T. Nakayama, *On Frobeniusean Algebras*, I, these Annals, Vol. 40 (1939), pp. 611–633; T. Nakayama, *Some Studies on Regular Representations, Induced Representations, and Modular Representations*, these Annals, Vol. 39 (1938), pp. 361–369; T. Nakayama and C. Nesbitt, *Note on Symmetric Algebras*, these Annals, Vol. 39 (1938), pp. 659–668; R. Brauer and C. Nesbitt, *On the Regular Representations of Algebras*, Proc. Nat. Acad. of Sci., Vol. 23 (1937), pp. 236–240; C. Nesbitt, *On the Regular Representations of Algebras*, these Annals, Vol. 39 (1938), pp. 634–658. We shall refer to the last two as B.N. and N.R., respectively.

[2] Cf. N.R., p. 639.

where $\mathfrak{N}$ is the radical of $\mathfrak{A}$ and

$$(2) \qquad \mathfrak{N} = \begin{pmatrix} 0 & & & \\ \mathfrak{C}_{21} & 0 & & \\ \vdots & \ddots & \ddots & \\ \mathfrak{C}_{t1} & \mathfrak{C}_{t2} \cdots & \cdots & 0 \end{pmatrix}, \qquad \mathfrak{A}^* = \begin{pmatrix} \mathfrak{C}_{11} & & & \\ 0 & \mathfrak{C}_{22} & & \\ \vdots & \ddots & \ddots & \\ 0 & \cdots & \cdots & \mathfrak{C}_{tt} \end{pmatrix}.$$

Let $F_1, F_2, \cdots, F_k$ denote the totality of distinct irreducible representations of $\mathfrak{A}$. Then each $\mathfrak{C}_{ii}$ is equivalent to one of the F_κ. It is well known that up to equivalence the irreducible constituents $\mathfrak{C}_{ii}$ of $\mathfrak{A}$ are uniquely determined.

As a part of $\mathfrak{A}$, $\mathfrak{C}_{ij}$ forms an additive group or module of matrices upon which $\mathfrak{A}$, itself considered as a module, is homomorphically mapped. We wish to consider $\mathfrak{C}_{ij}$ as a matrix module with $\mathfrak{A}$ as both a left and right operator system. For a matrix A of $\mathfrak{A}$, let us use the notation $C_{ij}(A)$, $(j \leqq i, i = 1, 2, \cdots, t)$ to denote the parts of A,

$$(3) \qquad A = \begin{pmatrix} C_{11}(A) & & & \\ C_{21}(A) & C_{22}(A) & & \\ \vdots & \vdots & \ddots & \\ C_{t1}(A) & C_{t2}(A) \cdots & \cdots & C_{tt}(A) \end{pmatrix}.$$

Let B be any element of $\mathfrak{A}$, and let B^* be the component of B in the semisimple subalgebra $\mathfrak{A}^*$. We define B as a left and as a right operator of $C_{ij}(A)$ by the relations below, using $\circ$ to distinguish this operation from ordinary matrix multiplication

$$(4) \qquad \begin{aligned} B \circ C_{ij}(A) &= C_{ii}(B) \cdot C_{ij}(A) = C_{ij}(B^* A), \\ C_{ij}(A) \circ B &= C_{ij}(A) \cdot C_{jj}(B) = C_{ij}(A B^*). \end{aligned}$$

Let us call a module which has $\mathfrak{A}$ as both right and left operator system an $(\mathfrak{A}, \mathfrak{A})$ module, and a homomorphism which is an operator mapping under the system of operators $\mathfrak{A}$, acting on both sides, an $(\mathfrak{A}, \mathfrak{A})$ homomorphism. Under definition (4), $\mathfrak{C}_{ij}$ is a simple $(\mathfrak{A}, \mathfrak{A})$ module. Moreover, each $\mathfrak{C}_{ij}$ is either a 0-part or there exist elements in $\mathfrak{A}$ such that the corresponding C_{ij} have any arbitrary components from K.

Accordingly, we shall call the $\mathfrak{C}_{ij}$ the *simple parts* of $\mathfrak{A}$. The simple parts $\mathfrak{C}_{ii}$, or irreducible constituents of $\mathfrak{A}$, have been rather thoroughly studied. Our first aim, then, is to develop a theory of simple parts which includes the $\mathfrak{C}_{ij}$, with $i \neq j$, that is, the simple parts of $\mathfrak{A}$ which appear in the radical $\mathfrak{N}$ of $\mathfrak{A}$.

A first remark is that, when $\mathfrak{A}$ is considered as an $(\mathfrak{A}, \mathfrak{A})$ module, it can be proved in a direct manner that each $\mathfrak{C}_{ij}$ which is not a 0-part is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to a composition factor group of $\mathfrak{A}$. This result may also be obtained by applying a useful theorem given recently by R. Brauer.[3]

[3] R. Brauer, op. cit.

When $\mathfrak{C}_{ii}$, $\mathfrak{C}_{jj}$ are respectively the irreducible constituents F_κ, F_λ of $\mathfrak{A}$, we shall say that $\mathfrak{C}_{ij}$ is of type (κ, λ). If $\kappa \neq \lambda$, we shall say that $\mathfrak{C}_{ij}$ is of mixed type, and if $\kappa = \lambda$ that $\mathfrak{C}_{ij}$ is of unmixed type. An $(\mathfrak{A}, \mathfrak{A})$ isomorphism may be defined for any two $\mathfrak{C}_{ij}$ of type (κ, λ) and, as a consequence, the corresponding factor groups of $\mathfrak{A}$ are also isomorphic. It is easy to see that two simple parts not of the same type are not $(\mathfrak{A}, \mathfrak{A})$ isomorphic. A representation $\bar{\mathfrak{A}}$ of $\mathfrak{A}$ may be considered as an $(\mathfrak{A}, \mathfrak{A})$ module. Again applying Brauer's Theorem it follows that the simple parts of type (κ, λ) of a representation $\bar{\mathfrak{A}}$, which is in reduced form, are $(\mathfrak{A}, \mathfrak{A})$ isomorphic to the simple parts of $\mathfrak{A}$ of type (κ, λ).

In N.R. extensive use was made of a basis system of $\mathfrak{A}$ first employed by Cartan[4], and which was there called the Cartan basis system. The coefficients in the expressions for the elements A of $\mathfrak{A}$ as linear combinations of the Cartan basis elements yield a set of matrix modules $\mathfrak{H}^\mu$, $(\mu = 1, 2, \cdots, m)$ upon which $\mathfrak{A}$, as a module, is homomorphically mapped. The $\mathfrak{H}^\mu$ may also be considered as $(\mathfrak{A}, \mathfrak{A})$ modules. We shall call these $\mathfrak{H}^\mu$ *elementary* (*related*) *modules* of $\mathfrak{A}$. The number m of elementary modules is equal to the composition length of $\mathfrak{A}$ considered as an $(\mathfrak{A}, \mathfrak{A})$ module, and each $\mathfrak{H}^\mu$ is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to a composition factor group of $\mathfrak{A}$. Conversely, to each composition factor group of a composition series of $\mathfrak{A}$ there corresponds an isomorphic elementary module. It follows that up to $(\mathfrak{A}, \mathfrak{A})$ isomorphism the set of elementary modules is uniquely determined.

We say that an elementary module $\mathfrak{H}^\mu$ is of type (κ, λ) if $\mathfrak{H}^\mu$ is defined by Cartan basis elements of type (κ, λ). Simple parts and elementary modules may also be classified by the powers of the radical to which they belong. A simple part $\mathfrak{C}_{ij}$ (an elementary module $\mathfrak{H}^\mu$) *belongs to* $\mathfrak{N}^h$ if this is the highest power of the radical such that all its elements are not mapped on the 0-matrix in $\mathfrak{C}_{ij}$ $(\mathfrak{H}^\mu)$. The principal relation between simple parts and elementary modules is that every simple part $\mathfrak{C}_{ij}$ of type (κ, λ) which belongs to $\mathfrak{N}^h$ is expressible as a linear combination of elementary modules $\mathfrak{H}^\mu$ of type (κ, λ) which belong to $\mathfrak{N}^j$, $j \leqq h$.

The form of a center element of $\mathfrak{A}$ can be computed. For a center element Z, $C_{ij}(Z) = 0$ if $\mathfrak{C}_{ij}$ is a simple part of mixed type, and $C_{ij}(Z)$ is of form $(c \cdot \delta_{mn})$, c and element of K, if $\mathfrak{C}_{ij}$ is of unmixed type. Let $c_{\kappa\lambda}$ denote, as in B.N., the Cartan Invariants. Then the rank ρ of the center is at most equal to the number $s = \sum_{\kappa=1}^{k} c_{\kappa\kappa}$ of elementary modules of unmixed type.

As an application of these concepts we obtain a decomposition of the matrix algebra $\mathfrak{A}$. By use of the concept of type, the simple parts of $\mathfrak{A}$ may be classified into 'blocks'. If $\mathfrak{C}_{ii}$, $\mathfrak{C}_{jj}$, $j < i$, do not belong to the same block then $\mathfrak{C}_{ij}$ must be a 0-part. An immediate consequence is that $\mathfrak{A}$ may be decomposed into parts which contain all simple parts belonging to a block. Each such part defines an invariant subalgebra which may not be decomposed into a direct sum of invariant subalgebras.

---

[4] E. Cartan, *Les groupes bilinearies et les systemes de nombres complexes*, Annales de Toulouse, 12B (1898), p. 1.

1

We consider a matrix algebra $\mathfrak{A}$, with unit element E, over an algebraically closed field K. We may assume that $\mathfrak{A}$ has no 0-constituent.[5] We take $\mathfrak{A}$ in the reduced form (1) and define operators for the parts $\mathfrak{C}_{ij}$ of $\mathfrak{A}$ by (4). The field K may be considered as included in the operator system by identifying the element k of K with the element $k \cdot E$ of $\mathfrak{A}$.

By means of (4) the parts $\mathfrak{C}_{ij}$ of $\mathfrak{A}$ (cf. (1)) may be regarded as $(\mathfrak{A}, \mathfrak{A})$ modules. If $\mathfrak{C}_{ij}$ contains only the 0-matrix, we call $\mathfrak{C}_{ij}$ a 0-part, and obviously, in this case, $\mathfrak{C}_{ij}$ is a simple module. Suppose, instead, that there exists an element A of $\mathfrak{A}$ such that $C_{ij}(A) \neq 0$. It is well known that elements L of $\mathfrak{A}$ can be chosen so that the corresponding $C_{ii}(L)$, $C_{jj}(L)$ have any components arbitrarily chosen from K. It follows from (4) that by suitably choosing right and left operators $L, M, \cdots$ to be applied to $C_{ij}(A)$ one obtains a system of matrices

$$C_{ij}(L^*AM^*) = C_{ii}(L)C_{ij}(A)C_{jj}(M)$$

in $\mathfrak{C}_{ij}$, from which may be generated matrices of $\mathfrak{C}_{ij}$ with arbitrary coefficients. Thus, the admissible subgroup which contains $C_{ij}(A)$ is the whole module C_{ij}. Let us call a module of matrices, each of m rows and n columns and with coefficients in K, a *complete module* if it contains all such matrices. We then have

THEOREM 1: *Each $\mathfrak{C}_{ij}$ is either a 0-part or a complete module. In either case $\mathfrak{C}_{ij}$ is a simple $(\mathfrak{A}, \mathfrak{A})$ module.*

We shall let $\mathfrak{N}^r$ denote the set which contains all elements of the form $\sum N_{\mu_1}N_{\mu_2}\cdots N_{\mu_r}$, where the N_{μ_i} are elements contained in the radical $\mathfrak{N}$. For a sufficiently large value of t, $\mathfrak{N}^t = 0$. These $\mathfrak{N}^r$ form invariant subalgebras of $\mathfrak{A}$. If we denote $\mathfrak{A}$ by $\mathfrak{N}^0$, we have

$$\mathfrak{A} = \mathfrak{N}^0 \supset \mathfrak{N}^1 \supset \mathfrak{N}^2 \supset \mathfrak{N}^3 \supset \cdots \supset \mathfrak{N}^{l-1} \supset \mathfrak{N}^l = 0.$$

We say that an element A belongs to $\mathfrak{N}^h$ if h is the largest value such that $\mathfrak{N}^h$ contains A. In particular, the elements of $\mathfrak{A}^*$ belong to $\mathfrak{N}^0$. The product of an element belonging to $\mathfrak{N}^h$ with an element belonging to $\mathfrak{N}^j$ belongs to $\mathfrak{N}^k$ with $k \geqq h + j$.

We can now prove

THEOREM 2: *A non-zero simple part $\mathfrak{C}_{ij}$ of $\mathfrak{A}$ is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to a composition factor group of $\mathfrak{A}$ itself considered as an $(\mathfrak{A}, \mathfrak{A})$ module.*

PROOF: Take the series

$$\mathfrak{A} \supset \mathfrak{N}^1 \supset \mathfrak{N}^2 \supset \mathfrak{N}^3 \supset \cdots \supset \mathfrak{N}^{l-1} \supset \mathfrak{N}^l = 0$$

and refine it to obtain a composition series

$$\mathfrak{A} \supset \cdots \supset \mathfrak{A}_{q-1} \supset \mathfrak{A}_q \supset \cdots \supset \mathfrak{A}_{n-1} \supset 0$$

---

[5] For, from conditions stated for $\mathfrak{A}$, we may show that if $\mathfrak{A}$ has 0-constituents, then it may be decomposed in the form $\mathfrak{A} = \begin{pmatrix} \mathfrak{A}_1 & 0 \\ 0 & 0 \end{pmatrix}$, where $\mathfrak{A}_1$ does not have 0-constituents, and $\mathfrak{A}$ is isomorphic to $\mathfrak{A}_1$.

of $\mathfrak{A}$. Let $\mathfrak{A}_q$ be the first group in this series such that for each of its elements A, $C_{ij}(A) = 0$. Then denoting elements of $\mathfrak{A}_{q-1}$ by A_{q-1}, we have that

$$A_{q-1} \to C_{ij}(A_{q-1})$$

is an $(\mathfrak{A}, \mathfrak{A})$ homomorphism, since for any element of $\mathfrak{A}$,

$$\begin{aligned} B \cdot A_{q-1} &= B^* \cdot A_{q-1} + N \cdot A_{q-1} \\ &\to C_{ij}(B^* A_{q-1}) + C_{ij}(N \cdot A_{q-1}). \end{aligned}$$

But $N \cdot A_{q-1} \subset \mathfrak{A}_q$ so that $C_{ij}(NA_{q-1}) = 0$ and, therefore,

$$B \cdot A_{q-1} \to C_{ij}(B^* A_{q-1}) = B \circ C_{ij}(A_{q-1}).$$

The same argument holds for right side multiplication. The kernel of this mapping (that is, the elements mapping into the identity element) contains $\mathfrak{A}_q$ and, therefore, the kernel is $\mathfrak{A}_q$. It follows that $\mathfrak{A}_{q-1}/\mathfrak{A}_q$ is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to the simple part $\mathfrak{C}_{ij}$.

We have said that if $\mathfrak{C}_{ii}$, $\mathfrak{C}_{jj}$ are respectively the irreducible constituents F_κ, F_λ of $\mathfrak{A}$, then the simple part $\mathfrak{C}_{ij}$ is of type (κ, λ). We now show that

THEOREM 3: *The non-zero simple parts $\mathfrak{C}_{ij}$ of a given type (κ, λ) are mutually isomorphic in the $(\mathfrak{A}, \mathfrak{A})$ sense. Simple parts of different types are not isomorphic.*

PROOF: It should be noted here that the word *isomorphic* is to have a meaning differing from that of Theorem 2. In Theorem 2, the isomorphism is obtained from $A \to C_{ij}(A)$, that is, $\mathfrak{C}_{ij}$ is considered as a set of matrices related to $\mathfrak{A}$. Here, $\mathfrak{C}_{ij}$ is just to represent the system of matrices.

Let $\mathfrak{C}_{ij}$, $\mathfrak{C}_{mn}$ be two simple parts of type (κ, λ) which are not 0-parts. An $(\mathfrak{A}, \mathfrak{A})$ mapping of $\mathfrak{C}_{ij}$ upon $\mathfrak{C}_{mn}$ is obtained by the relation $C_{ij} \to C_{mn}$ if $C_{ij} = C_{mn}$ where C_{ij}, C_{mn} denote matrices of $\mathfrak{C}_{ij}$, $\mathfrak{C}_{mn}$, respectively. For if B is any element of $\mathfrak{A}$, then

$$\begin{aligned} B \circ C_{ij} &= C_{ii}(B) \cdot C_{ij} = F_\kappa(B) \cdot C_{ij} \\ &= C_{mm}(B) \cdot C_{mn} = B \circ C_{mn}. \end{aligned}$$

On the other hand, if $\mathfrak{C}_{ij}$, $\mathfrak{C}_{mn}$ are of types (κ, λ), (μ, ν), respectively, with $\kappa \neq \mu$, then any relation $C_{ij} \to C_{mn}$ is not an $(\mathfrak{A}, \mathfrak{A})$ operator mapping, for we may choose B in $\mathfrak{A}$ such that $C_{ii}(B) = F_\kappa(B) = 0$ while $C_{mm}(B) = F_\mu(B) = E_{f_\mu}$ where E_{f_μ} is a unit matrix of proper degree.

2

The basis of $\mathfrak{A}$ which seems best adapted to the study of simple parts is a basis system first employed by Cartan.[6] This basis was used by C. Nesbitt in his study of regular representations of linear associative algebras.[7]

We denote, as before, the irreducible representations of $\mathfrak{A}$ by F_1, F_2, F_3,

[6] E. Cartan, op. cit.

[7] A detailed description of the Cartan basis system is given in N.R. section 2.

$\cdots$, F_k, and their degrees by $f_1, f_2, \cdots, f_k$. Let $E_\kappa(m, n)$ be the matrix of $\mathfrak{A}$ which has 1 for the (m, n) component of each $\mathfrak{C}_{ii}$ which is equal to F_κ, and has zero elsewhere. The relations

$$E_\mu(m, n)\cdot E_\nu(p, q) = \begin{cases} 0 & \text{if either } \mu \neq \nu, \text{ or } p \neq q, \\ E_\mu(m, q) & \text{if } \mu = \nu, n = p, \end{cases}$$

follow from the definition of the $E_\mu(m, n)$.

Let $E_\kappa = \sum_i E_\kappa(ii)$. Then the E_κ, $(\kappa = 1, 2, \cdots, k)$, form a system of mutually idempotent elements. In fact, E_κ is the unit element of a certain simple algebra which is a summand in a decomposition of $\mathfrak{A}^*$ into a direct sum of simple algebras, and $\sum_{\kappa=1}^{k} E_\kappa = E$, where E is the unit element of $\mathfrak{A}$.

A matrix A of $\mathfrak{A}$ is said to be of type (κ, λ) if $E_\kappa \cdot A \cdot E_\lambda = A$. In particular, $E_\kappa(m, n)$ is of type (κ, κ). An element of type (κ, λ), $\kappa \neq \lambda$, we shall say is of mixed type, and if $\kappa = \lambda$, of unmixed type. The product of an element of type (κ, λ) and an element of type (μ, ν) is an element of type (κ, ν). The product of an element of type (κ, λ) and of an element of type (μ, ν) vanishes for $\lambda \neq \mu$. A matrix A of type (κ, λ) has non-zero components only in simple parts $\mathfrak{C}_{ij}$ of type (κ, λ).

A basis set $B^1_{\kappa\lambda}, B^2_{\kappa\lambda}, \cdots, B^t_{\kappa\lambda}$ is first obtained for those matrices of $\mathfrak{A}$ of type (κ, λ) which have non-zero components only in the upper left corners of simple parts of type (κ, λ). The rank t of this set is the Cartan Invariant, $c_{\kappa\lambda}$.[8] A basis for the whole algebra is then given by the set

$$E_\kappa(a1)B^\mu_{\kappa\lambda}E_\lambda(1b),$$
$$\mu = 1, 2, \cdots, c_{\kappa\lambda}, \qquad a = 1, 2, \cdots, f_\kappa,$$
$$b = 1, 2, \cdots, f_\lambda, \qquad \kappa, \lambda = 1, 2, \cdots, k.$$

This basis can be so chosen that any element A of $\mathfrak{A}$ which belongs to $\mathfrak{N}^\tau$ is expressible in terms of basis elements which belong to $\mathfrak{N}^\rho$, for $\rho \geqq \tau$.[9]

For convenience we change slightly the above notation. We take together in one set all the elements $B^\mu_{\kappa\lambda}$ for all κ, λ, and μ and redefine the superscript μ to enumerate these elements, independently of type, in such a way as to take first all those elements which belong to $\mathfrak{N}^0$, namely $E_1(11), \cdots, E_k(11)$, then next those which belong to $\mathfrak{N}^1$, and so on. μ will now have the range $1, 2, \cdots, m$ where $m = \sum_{\kappa,\lambda=1}^{k} c_{\kappa\lambda}$. Where the type of the element need not be explicitly stated we shall drop the subscripts κ, λ from the symbols $B^\mu_{\kappa\lambda}$.

3

Expressing the element A of $\mathfrak{A}$ as a linear combination of the elements of the Cartan basis system, we have

$$A = \sum_{a,b,\mu} h^\mu_{ab}(A)E_\kappa(a1)B^\mu_{\kappa\lambda}E_\lambda(1b). \tag{5}$$

[8] See B.N. or N.R.

[9] See N.R., p. 641.

Here each $B^{\mu}_{\kappa\lambda}$ is a basis element having components different from zero only in the upper left hand corners of simple parts $\mathfrak{C}_{ij}$ which are of type (κ, λ).

We denote the matrix $(h^{\mu}_{ab}(A))_{ab}$ (cf. (5)) by $H^{\mu}(A)$. The module consisting of the matrices $H^{\mu}(A)$ will be denoted by $\mathfrak{H}^{\mu}$. We shall call $\mathfrak{H}^{\mu}$ an *elementary (related) module* of $\mathfrak{A}$. $\mathfrak{H}^{\mu}$ is evidently a complete module, since the coefficients in (5) may be taken arbitrarily from K.

An elementary module will be said to be of type (κ, λ) if it results from, or is defined by, Cartan basis elements of type (κ, λ). We shall write $\mathfrak{H}^{\mu}_{\kappa\lambda}$ to denote that $\mathfrak{H}^{\mu}$ is of type (κ, λ). If $\kappa \neq \lambda$ we shall call $\mathfrak{H}^{\mu}_{\kappa\lambda}$ an elementary module of *mixed type*, and if $\kappa = \lambda$ we shall call $\mathfrak{H}^{\mu}_{\kappa\lambda}$ an elementary module of *unmixed type*.

Let us say that a module $\mathfrak{M}$ of matrices is related to $\mathfrak{N}^{\rho}$ if $\mathfrak{N}^{\rho}$ is homomorphically mapped on $\mathfrak{M}$. In general, we shall deal with modules $\mathfrak{M}$ which are related to $\mathfrak{N}^{0} = \mathfrak{A}$, but it is convenient sometimes to consider $\mathfrak{M}$ as the map of only the subalgebra $\mathfrak{N}^{\rho}$ rather than $\mathfrak{A}$ itself (see p. 157). A related module $\mathfrak{M}$ will be said to belong to $\mathfrak{N}^{\sigma}$ if σ is the largest value such that for at least one element N_{σ} belonging to $\mathfrak{N}^{\sigma}$, the corresponding matrix $M(N_{\sigma})$ of $\mathfrak{M}$ is not zero. In particular, the elementary module $\mathfrak{H}^{\mu}$ is related to $\mathfrak{A}$, and belongs to the same power of $\mathfrak{N}$ as does the corresponding basis element B^{μ}.

LEMMA 1: *Let $\mathfrak{M}$ be a module for which the concept of type is defined. If*

i). *$\mathfrak{M}$ is of type (κ, λ), and its elements are $f_{\kappa} \times f_{\lambda}$ matrices,*

ii). *$\mathfrak{M}$ is related to $\mathfrak{N}^{\rho}$, and belongs to $\mathfrak{N}^{\sigma}$, and*

iii). *$\mathfrak{M}$ is a simple $(\mathfrak{A}, \mathfrak{A})$ module such that for any element B of $\mathfrak{A}$, and element N_{ρ} of $\mathfrak{N}^{\rho}$*

$$
\begin{aligned}
B \circ M(N_{\rho}) &= F_{\kappa}(B) \cdot M(N_{\rho}) = M(B^{*}N_{\rho}) \\
M(N_{\rho}) \circ B &= M(N_{\rho}) \cdot F_{\lambda}(B) = M(N_{\rho}B^{*}),
\end{aligned} \tag{6}
$$

then it follows that as a module related to $\mathfrak{N}^{\rho}$, $M(N_{\rho}) = \sum_{\mu} m_{\mu} H^{\mu}_{\kappa\lambda}(N_{\rho})$, where m_{μ} are fixed scalars, N_{ρ} is any element of $\mathfrak{N}^{\rho}$, and where the elementary module $\mathfrak{H}^{\mu}_{\kappa\lambda}$ belongs to $\mathfrak{N}^{\tau}$, $\rho \leqq \tau \leqq \sigma$, and is considered as related to $\mathfrak{N}^{\sigma}$.

PROOF: We shall use the notation $(U_{rs})_{pq}$ to denote a matrix with 1 as its (r, s) component, and with zero elsewhere. Then operating with $E_{\kappa}(1, 1)$, $E_{\lambda}(1, 1)$ on $M(B^{\mu}_{\kappa\lambda}) = (m^{\mu}_{pq})$, we obtain

$$
\begin{aligned}
E_{\kappa}(11) \circ M(B^{\mu}_{\kappa\lambda}) \circ E_{\lambda}(11) &= (U_{11})_{rs} \cdot (m^{\mu}_{pq}) \cdot (U_{11})_{ut} = (m^{\mu}_{11}) \cdot (U_{11})_{pq} \\
&= M(E_{\kappa}(11) B^{\mu}_{\kappa\lambda} E_{\lambda}(11)) = M(B^{\mu}_{\kappa\lambda}).
\end{aligned}
$$

Dropping the subscripts 11 we have

$$
M(B^{\mu}_{\kappa\lambda}) = (m^{\mu} U_{11})_{pq} .
$$

Here $m^{\mu} \neq 0$ only if $B^{\mu}_{\kappa\lambda}$ belongs to $\mathfrak{N}^{\tau}$, $\tau \leqq \sigma$. Also

$$
M(E_{\kappa}(a1) B^{\mu}_{\kappa\lambda} E_{\lambda}(1b)) = (m^{\mu} U_{ab})_{pq} .
$$

$\cdots$, F_k, and their degrees by $f_1, f_2, \cdots, f_k$. Let $E_\kappa(m, n)$ be the matrix of $\mathfrak{A}$ which has 1 for the (m, n) component of each $\mathfrak{C}_{ii}$ which is equal to F_κ, and has zero elsewhere. The relations

$$E_\mu(m, n) \cdot E_\nu(p, q) = \begin{cases} 0 & \text{if either } \mu \neq \nu, \text{ or } p \neq q, \\ E_\mu(m, q) & \text{if } \mu = \nu, n = p, \end{cases}$$

follow from the definition of the $E_\mu(m, n)$.

Let $E_\kappa = \sum_i E_\kappa(ii)$. Then the E_κ, $(\kappa = 1, 2, \cdots, k)$, form a system of mutually idempotent elements. In fact, E_κ is the unit element of a certain simple algebra which is a summand in a decomposition of $\mathfrak{A}^*$ into a direct sum of simple algebras, and $\sum_{\kappa=1}^{k} E_\kappa = E$, where E is the unit element of $\mathfrak{A}$.

A matrix A of $\mathfrak{A}$ is said to be of type (κ, λ) if $E_\kappa \cdot A \cdot E_\lambda = A$. In particular, $E_\kappa(m, n)$ is of type (κ, κ). An element of type (κ, λ), $\kappa \neq \lambda$, we shall say is of mixed type, and if $\kappa = \lambda$, of unmixed type. The product of an element of type (κ, λ) and an element of type (μ, ν) is an element of type (κ, ν). The product of an element of type (κ, λ) and of an element of type (μ, ν) vanishes for $\lambda \neq \mu$. A matrix A of type (κ, λ) has non-zero components only in simple parts $\mathfrak{C}_{ij}$ of type (κ, λ).

A basis set $B^1_{\kappa\lambda}, B^2_{\kappa\lambda}, \cdots, B^t_{\kappa\lambda}$ is first obtained for those matrices of $\mathfrak{A}$ of type (κ, λ) which have non-zero components only in the upper left corners of simple parts of type (κ, λ). The rank t of this set is the Cartan Invariant, $c_{\kappa\lambda}$.[8] A basis for the whole algebra is then given by the set

$$E_\kappa(a1) B^\mu_{\kappa\lambda} E_\lambda(1b),$$
$$\mu = 1, 2, \cdots, c_{\kappa\lambda}, \qquad a = 1, 2, \cdots, f_\kappa,$$
$$b = 1, 2, \cdots, f_\lambda, \qquad \kappa, \lambda = 1, 2, \cdots, k.$$

This basis can be so chosen that any element A of $\mathfrak{A}$ which belongs to $\mathfrak{N}^\tau$ is expressible in terms of basis elements which belong to $\mathfrak{N}^\rho$, for $\rho \geqq \tau$.[9]

For convenience we change slightly the above notation. We take together in one set all the elements $B^\mu_{\kappa\lambda}$ for all κ, λ, and μ and redefine the superscript μ to enumerate these elements, independently of type, in such a way as to take first all those elements which belong to $\mathfrak{N}^0$, namely $E_1(11), \cdots, E_k(11)$, then next those which belong to $\mathfrak{N}^1$, and so on. μ will now have the range $1, 2, \cdots, m$ where $m = \sum_{\kappa,\lambda=1}^{k} c_{\kappa\lambda}$. Where the type of the element need not be explicitly stated we shall drop the subscripts κ, λ from the symbols $B^\mu_{\kappa\lambda}$.

3

Expressing the element A of $\mathfrak{A}$ as a linear combination of the elements of the Cartan basis system, we have

$$A = \sum_{a,b,\mu} h^\mu_{ab}(A) E_\kappa(a1) B^\mu_{\kappa\lambda} E_\lambda(1b). \tag{5}$$

[8] See B.N. or N.R.

[9] See N.R., p. 641.

Here each $B^{\mu}_{\kappa\lambda}$ is a basis element having components different from zero only in the upper left hand corners of simple parts $\mathfrak{S}_{ij}$ which are of type (κ, λ).

We denote the matrix $(h^{\mu}_{ab}(A))_{ab}$ (cf. (5)) by $H^{\mu}(A)$. The module consisting of the matrices $H^{\mu}(A)$ will be denoted by $\mathfrak{H}^{\mu}$. We shall call $\mathfrak{H}^{\mu}$ an *elementary (related) module* of $\mathfrak{A}$. $\mathfrak{H}^{\mu}$ is evidently a complete module, since the coefficients in (5) may be taken arbitrarily from K.

An elementary module will be said to be of type (κ, λ) if it results from, or is defined by, Cartan basis elements of type (κ, λ). We shall write $\mathfrak{H}^{\mu}_{\kappa\lambda}$ to denote that $\mathfrak{H}^{\mu}$ is of type (κ, λ). If $\kappa \neq \lambda$ we shall call $\mathfrak{H}^{\mu}_{\kappa\lambda}$ an elementary module of *mixed type*, and if $\kappa = \lambda$ we shall call $\mathfrak{H}^{\mu}_{\kappa\lambda}$ an elementary module of *unmixed type*.

Let us say that a module $\mathfrak{M}$ of matrices is related to $\mathfrak{N}^{\rho}$ if $\mathfrak{N}^{\rho}$ is homomorphically mapped on $\mathfrak{M}$. In general, we shall deal with modules $\mathfrak{M}$ which are related to $\mathfrak{N}^{0} = \mathfrak{A}$, but it is convenient sometimes to consider $\mathfrak{M}$ as the map of only the subalgebra $\mathfrak{N}^{\rho}$ rather than $\mathfrak{A}$ itself (see p. 157). A related module $\mathfrak{M}$ will be said to belong to $\mathfrak{N}^{\sigma}$ if σ is the largest value such that for at least one element N_{σ} belonging to $\mathfrak{N}^{\sigma}$, the corresponding matrix $M(N_{\sigma})$ of $\mathfrak{M}$ is not zero. In particular, the elementary module $\mathfrak{H}^{\mu}$ is related to $\mathfrak{A}$, and belongs to the same power of $\mathfrak{N}$ as does the corresponding basis element B^{μ}.

LEMMA 1: *Let $\mathfrak{M}$ be a module for which the concept of type is defined. If*

i). *$\mathfrak{M}$ is of type (κ, λ), and its elements are $f_{\kappa} \times f_{\lambda}$ matrices,*

ii). *$\mathfrak{M}$ is related to $\mathfrak{N}^{\rho}$, and belongs to $\mathfrak{N}^{\sigma}$, and*

iii). *$\mathfrak{M}$ is a simple $(\mathfrak{A}, \mathfrak{A})$ module such that for any element B of $\mathfrak{A}$, and element N_{ρ} of $\mathfrak{N}^{\rho}$*

$$
\begin{aligned}
B \circ M(N_{\rho}) &= F_{\kappa}(B) \cdot M(N_{\rho}) = M(B^{*}N_{\rho}) \\
M(N_{\rho}) \circ B &= M(N_{\rho}) \cdot F_{\lambda}(B) = M(N_{\rho}B^{*}),
\end{aligned} \tag{6}
$$

then it follows that as a module related to $\mathfrak{N}^{\rho}$, $M(N_{\rho}) = \sum_{\mu} m_{\mu} H^{\mu}_{\kappa\lambda}(N_{\rho})$, where m_{μ} are fixed scalars, N_{ρ} is any element of $\mathfrak{N}^{\rho}$, and where the elementary module $\mathfrak{H}^{\mu}_{\kappa\lambda}$ belongs to $\mathfrak{N}^{\tau}$, $\rho \leqq \tau \leqq \sigma$, and is considered as related to $\mathfrak{N}^{\sigma}$.

PROOF: We shall use the notation $(U_{rs})_{pq}$ to denote a matrix with 1 as its (r, s) component, and with zero elsewhere. Then operating with $E_{\kappa}(1, 1)$, $E_{\lambda}(1, 1)$ on $M(B^{\mu}_{\kappa\lambda}) = (m^{\mu}_{pq})$, we obtain

$$
\begin{aligned}
E_{\kappa}(11) \circ M(B^{\mu}_{\kappa\lambda}) \circ E_{\lambda}(11) &= (U_{11})_{rs} \cdot (m^{\mu}_{pq}) \cdot (U_{11})_{ut} = (m^{\mu}_{11}) \cdot (U_{11})_{pq} \\
&= M(E_{\kappa}(11) B^{\mu}_{\kappa\lambda} E_{\lambda}(11)) = M(B^{\mu}_{\kappa\lambda}).
\end{aligned}
$$

Dropping the subscripts 11 we have

$$
M(B^{\mu}_{\kappa\lambda}) = (m^{\mu} U_{11})_{pq} .
$$

Here $m^{\mu} \neq 0$ only if $B^{\mu}_{\kappa\lambda}$ belongs to $\mathfrak{N}^{\tau}$, $\tau \leqq \sigma$. Also

$$
M(E_{\kappa}(a1) B^{\mu}_{\kappa\lambda} E_{\lambda}(1b)) = (m^{\mu} U_{ab})_{pq} .
$$

We have finally that

$$M(N_\rho) = M\Big(\sum_{\mu,a,b} h^\mu_{ab}(N_\rho) E_\kappa(a1) B^\mu_{\kappa\lambda} E_\lambda(1b)\Big)$$
$$= \sum_\mu m^\mu H^\mu_{\kappa\lambda}(N^\rho)$$

where $\mathfrak{H}^\mu_{\kappa\lambda}$ belongs to $\mathfrak{N}^\tau$, $\rho \leqq \tau \leqq \sigma$, since N_ρ is an element of $\mathfrak{N}^\rho$, and m belongs to $\mathfrak{N}^\sigma$.

Relating the simple parts of $\mathfrak{A}$ to the elementary modules of $\mathfrak{A}$, we have

THEOREM 4: *A simple part $\mathfrak{C}_{ij}$ of $\mathfrak{A}$ of type (κ, λ) which belongs to $\mathfrak{N}^r$ is expressible as a linear combination of the elementary modules of type (κ, λ) which belong to $\mathfrak{N}^r$, $\tau \leqq r$.*

PROOF: We consider $\mathfrak{C}_{ij}$ as a simple module related to $\mathfrak{A} = \mathfrak{N}^0$, and belonging to $\mathfrak{N}^r$, and apply the lemma. It follows that $\mathfrak{C}_{ij}$ is of the form $\sum_\mu d^\mu_{ij} H^\mu_{\kappa\lambda}$, that is, for each element A of $\mathfrak{A}$ we have that

$$C_{ij}(A) = \sum_\mu d^\mu_{ij} H^\mu_{\kappa\lambda}(A),$$

where each $H^\mu_{\kappa\lambda}(A)$ belongs to some N^τ, $\tau \leqq r$.

We now prove

THEOREM 5: *The number m of elementary modules is equal to the composition length of $\mathfrak{A}$ considered as an $(\mathfrak{A}, \mathfrak{A})$ module. Each elementary module $\mathfrak{H}$ is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to a composition factor group of $\mathfrak{A}$, and conversely, to each factor group of $\mathfrak{A}$ there corresponds an isomorphic elementary module.*

PROOF: We will construct a composition series for $\mathfrak{A}$ in terms of the Cartan basis elements. As indicated above, we take the whole set of elements $B^\mu_{\kappa\lambda}$ $(\kappa, \lambda = 1, 2, 3, \cdots, k)$ and use the superscript μ $(\mu = 1, 2, 3, \cdots, m)$ to enumerate these elements, taking first those elements which are not in $\mathfrak{N}^1$, then those which are in $\mathfrak{N}^1$ but not in $\mathfrak{N}^2$, and so on. Let $\mathfrak{A}_{s-1}$ denote the subgroup of $\mathfrak{A}$ consisting of all linear combinations with coefficients in K of the elements

$$(\kappa, \lambda = 1, 2, \cdots, \text{or } k),$$
$$(7) \qquad E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b), \qquad (a = 1, 2, \cdots, f_\kappa; \quad b = 1, 2, \cdots, f_\lambda),$$
$$(\tau = S, S+1, \cdots, m).$$

In particular, $\mathfrak{A}_0 = \mathfrak{A}$. We form the product

$$A \cdot E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b) = (A^* + N) \cdot E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b).$$

Here

$$A^* \cdot E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b) = \sum_{\rho,m,n} f^\rho_{mn}(A) E_\rho(m\,n) \cdot E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b)$$
$$= \sum_m f^\kappa_{ma}(A) E_\kappa(m1) B^\tau_{\kappa\lambda} E_\lambda(1b)$$

is again in $\mathfrak{A}_{s-1}$. So also is $N \cdot E_\kappa(a1) B^\tau_{\kappa\lambda} E_\lambda(1b)$, since this belongs to a higher

power of the radical than does $B^r_{\kappa\lambda}$ and is, therefore, expressible in terms of the elements (7). Then the whole product is in $\mathfrak{A}_{s-1}$. The same is true when A is a right factor, so we obtain that $\mathfrak{A}_{s-1}$ is an admissible subgroup of $\mathfrak{A}$. Moreover, the factor group $\mathfrak{A}_{s-1}/\mathfrak{A}_s$ consisting of the residue classes

$$\langle \sum_{a,b} \lambda^s_{ab} E_\kappa(a1) B^s_{\kappa\lambda} E_\lambda(1b) \rangle, \qquad \text{modulo } \mathfrak{A}_s,$$

is easily seen by similar calculations to be simple. It follows that

$$\mathfrak{A}_0 \supset \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_{m-1} \supset 0$$

is a composition series of $\mathfrak{A}$, and the first part of the theorem is proved.

Again, we may easily calculate that

$$\langle \sum_{a,b} h^s_{ab}(A) E_\kappa(a1) B^s_{\kappa\lambda} E_\lambda(1b) \rangle \to H^s_{\kappa\lambda}(A)$$

is an $(\mathfrak{A}, \mathfrak{A})$ isomorphic mapping of $\mathfrak{A}_{s-1}/\mathfrak{A}_s$ upon the elementary module $\mathfrak{H}^s_{\kappa\lambda}$ (cf. (7)). From this follow the other statements in the theorem.

It may be noted here that from Theorem 4 and Theorem 5 another proof of Theorem 2 may be obtained.

We can now prove the converse of Theorem 2

THEOREM 6: *To each composition factor group of $\mathfrak{A}$ there corresponds at least one non-zero simple part $\mathfrak{C}_{ij}$.*

PROOF: By Theorem 5 we have that to each composition factor group of $\mathfrak{A}$ there corresponds an isomorphic elementary module $\mathfrak{H}^\rho$, say of type (μ, ν). Then there exist non-zero simple parts of type (μ, ν), and by the same argument as that used in Theorem 3 it may be shown that each non-zero simple part of type (μ, ν) is $(\mathfrak{A}, \mathfrak{A})$ isomorphic to $\mathfrak{H}^\rho$.

4

A system of matrices $\bar{\mathfrak{A}}$ is called a representation of $\mathfrak{A}$ if $\mathfrak{A}$ is homomorphically mapped on $\bar{\mathfrak{A}}$. We may define $\mathfrak{A}$ as a left and right operator system for a representation $\bar{\mathfrak{A}}$ by the relations

$$\begin{aligned} A \cdot \bar{A}(A_1) &= \bar{A}(A \cdot A_1) = \bar{A}(A) \cdot \bar{A}(A_1), \\ \bar{A}(A_1) \cdot A &= \bar{A}(A_1 \cdot A) = \bar{A}(A_1) \cdot \bar{A}(A). \end{aligned} \tag{8}$$

Further, if the representation $\bar{\mathfrak{A}}$ is taken in reduced form, then a simple part $\bar{\mathfrak{C}}_{ij}$ of $\bar{\mathfrak{A}}$ may be considered as an $(\mathfrak{A}, \mathfrak{A})$ module if we define the elements of $\mathfrak{A}$ as operators in $\bar{\mathfrak{C}}_{ij}$ by relations of the form (4). We then obtain

THEOREM 7: *The simple parts $\bar{\mathfrak{C}}_{ij}$, of type (κ, λ), of a representation $\bar{\mathfrak{A}}$ of $\mathfrak{A}$ are $(\mathfrak{A}, \mathfrak{A})$ isomorphic[10] to simple parts $\mathfrak{C}_{mn}$, of type (κ, λ), of $\mathfrak{A}$.*

[10] As in Theorem 3, the word *isomorphism* here has a somewhat different meaning from that of Theorem 2.

PROOF: By Theorem 2 we have that each simple part $\bar{\mathfrak{C}}_{ij}$ of $\bar{\mathfrak{A}}$ is isomorphic to a composition factor group of a composition series for $\bar{\mathfrak{A}}$. Let

$$\bar{\mathfrak{A}} \supset \bar{\mathfrak{A}}_1 \supset \bar{\mathfrak{A}}_2 \supset \cdots \supset \bar{\mathfrak{A}}_l = 0, \tag{9}$$

$$\mathfrak{A} \supset \mathfrak{A}_1 \supset \mathfrak{A}_2 \supset \cdots \supset \mathfrak{A}_n = 0, \tag{10}$$

be composition series for $\bar{\mathfrak{A}}$, $\mathfrak{A}$, respectively, and suppose

$$\bar{\mathfrak{C}}_{ij} \cong \bar{\mathfrak{A}}_{p-1}/\bar{\mathfrak{A}}_p .$$

$\bar{\mathfrak{A}}$ is an $(\mathfrak{A}, \mathfrak{A})$ module upon which $\mathfrak{A}$ is homomorphically mapped by the relation $A \to \bar{A}(A)$. Then the Jordan-Hölder Theorem gives us that to the factor group $\bar{\mathfrak{A}}_{p-1}/\bar{\mathfrak{A}}_p$ of (9) there corresponds a factor group $\mathfrak{A}_{q-1}/\mathfrak{A}_q$ of the series (10) such that

$$\bar{\mathfrak{A}}_{p-1}/\bar{\mathfrak{A}}_p \cong \mathfrak{A}_{q-1}/\mathfrak{A}_q .$$

In addition we have from Theorem 6 that to the factor group $\mathfrak{A}_{q-1}/\mathfrak{A}_q$ there is at least one non-zero isomorphic simple part $\mathfrak{C}_{mn}$ of $\mathfrak{A}$,

$$\mathfrak{A}_{q-1}/\mathfrak{A}_q \cong \mathfrak{C}_{mn} .$$

Combining these relations, and observing that all these isomorphisms are $(\mathfrak{A}, \mathfrak{A})$ isomorphisms, we obtain

$$\bar{\mathfrak{C}}_{ij} \cong \mathfrak{C}_{mn}$$

with operators $(\mathfrak{A}, \mathfrak{A})$. This completes the proof of the statement.

Now suppose that $\bar{\mathfrak{A}}$ is a representation of $\mathfrak{A}$, and that $\bar{\mathfrak{A}}$ is in reduced form. Let us denote the element of $\bar{\mathfrak{A}}$ corresponding to the element A of $\mathfrak{A}$ by $\bar{A}$, $A \to \bar{A}$; further, we shall consider $\bar{\mathfrak{A}}$ as a direct sum of its radical $\bar{\mathfrak{A}}_R$ and a semi-simple subalgebra, $\bar{\mathfrak{A}}_S$, $\bar{\mathfrak{A}} = \bar{\mathfrak{A}}_S + \bar{\mathfrak{A}}_R$, where $\bar{\mathfrak{A}}_S$ is completely decomposed into its irreducible constituents. We use $\bar{A}_S$, $\bar{A}_R$ to denote the components in $\bar{\mathfrak{A}}_S$, $\bar{\mathfrak{A}}_R$ of the element $\bar{A}$ of $\bar{\mathfrak{A}}$, $\bar{A} = \bar{A}_S + \bar{A}_R$.

The set $\bar{\mathfrak{A}}^*$ of elements in the representation $\bar{\mathfrak{A}}$ which correspond to the semi-simple subalgebra $\mathfrak{A}^*$ of $\mathfrak{A}$ form a semisimple subalgebra of $\bar{\mathfrak{A}}$ equivalent to $\bar{\mathfrak{A}}_S$. Since $\bar{\mathfrak{A}}^*, \bar{\mathfrak{A}}_S$ have the same irreducible constituents, they are equivalent algebras, but are not necessarily identical. It is easy to construct examples where $\bar{\mathfrak{A}}^*$ is not $\bar{\mathfrak{A}}_S$.

Let us suppose $\bar{\mathfrak{H}}_1, \cdots, \bar{\mathfrak{H}}_g$ are a set of elementary modules for the algebra $\bar{\mathfrak{A}}$. We wish to give explicitly the relation between the modules $\bar{\mathfrak{H}}$ of $\bar{\mathfrak{A}}$, and the elementary modules $\mathfrak{H}$ of $\mathfrak{A}$.

The irreducible constituents of $\bar{\mathfrak{A}}$ are equivalent to irreducible constituents of $\mathfrak{A}$. For simplicity, let us suppose they are identical with constituents of $\mathfrak{A}$. Let $\bar{\mathfrak{H}}$ be of type (κ, λ) and belong to the σ^{th} power of the radical of $\bar{\mathfrak{A}}$. We may consider $\bar{\mathfrak{H}}$ as a module related to $\mathfrak{A}$ by setting $\bar{H}(A) = \bar{H}(\bar{A})$. As such, $\bar{\mathfrak{H}}$ belongs to $\mathfrak{N}^\sigma$. Let B be any element of $\mathfrak{A}$, and N_σ an element of $\mathfrak{N}^\sigma$. Then

$$F_\kappa(B)\bar{H}(N_\sigma) = F_\kappa(\bar{B})\bar{H}(\bar{N}_\sigma) = \bar{H}(\bar{B}_S\bar{N}_\sigma).$$

But $\bar{B}_S \equiv \bar{B}^*$, mod $\bar{\mathfrak{A}}_R$, so that $\bar{B}_S\bar{N}_\sigma \equiv \bar{B}^*\bar{N}_\sigma$, mod $\bar{\mathfrak{A}}_R^{\sigma+1}$, and as $\bar{\mathfrak{H}}$ belongs to the σ^{th} power of the radical $\bar{\mathfrak{A}}_R$ of $\bar{\mathfrak{A}}$, we have

$$\bar{H}(\bar{B}_S\bar{N}_\sigma) = \bar{H}(\bar{B}^*\bar{N}_\sigma) = \bar{H}(\overline{B^*N_\sigma}) = \bar{H}(B^*N_\sigma).$$

We may then view $\bar{\mathfrak{H}}$ as an $(\mathfrak{A}, \mathfrak{A})$ module of type (κ, λ), related and belonging to $\mathfrak{N}^\sigma$, such that

$$B\circ\bar{H}(N_\sigma) = F_\kappa(B)\bar{H}(N_\sigma) = \bar{H}(B^*N_\sigma)$$

with a similar relation for right operators. Applying Lemma 1, we have that as a module related to $\mathfrak{N}^\sigma$, $\bar{\mathfrak{H}}$ is a linear combination of the elementary modules $\mathfrak{H}^\mu_{\kappa\lambda}$ of type (κ, λ) which belong to $\mathfrak{N}^\sigma$.[11] It follows, also, that a simple part $\bar{\mathfrak{C}}$ of $\bar{\mathfrak{A}}$ of type (κ, λ) which belongs to the σ^{th} power of the radical of $\bar{\mathfrak{A}}$ may, considered as a module related to $\mathfrak{N}^\sigma$, be expressed as a linear combination of the elementary modules $\mathfrak{H}^\mu_{\kappa\lambda}$ of type (κ, λ) which belong to $\mathfrak{N}^\sigma$.

If $\bar{\mathfrak{A}}^* = \bar{\mathfrak{A}}_S$ then the relation

$$F_\kappa(B)\bar{H}(A) = \bar{H}(B^*A)$$

holds for all elements A of $\mathfrak{A}$. Again applying Lemma 1 we obtain that $\bar{\mathfrak{H}}$ considered as a module related to $\mathfrak{A}$ itself is expressible as a linear combination of the elementary modules $\mathfrak{H}$ of type (κ, λ) which belong to $\mathfrak{N}^\tau, \tau \leqq \sigma$.

If $\bar{\mathfrak{A}}^*$ is not identical to $\bar{\mathfrak{A}}_S$, that is, if $\bar{\mathfrak{A}}^*$ is not completely decomposed, then we can go over to an equivalent representation $\bar{\mathfrak{A}}$, such that $\bar{\mathfrak{A}}^* = \bar{\mathfrak{A}}_S$ where $\bar{\mathfrak{A}}^*$, $\bar{\mathfrak{A}}_S$ have the same meaning with regard to $\bar{\mathfrak{A}}$ as do $\bar{\mathfrak{A}}^*$ and $\bar{\mathfrak{A}}_S$ in regard to $\bar{\mathfrak{A}}$. We may then state:

THEOREM 8: *To any representation of $\mathfrak{A}$ there exists an equivalent representation whose simple parts, considered as modules related to $\mathfrak{A}$, are expressible as linear combinations of the elementary modules $\mathfrak{H}$ of $\mathfrak{A}$.*

5

It is easy now to discuss the center of $\mathfrak{A}$. We note that

THEOREM 9: *For a center element Z of $\mathfrak{A}$, $C_{ij}(Z) = 0$ if $\mathfrak{C}_{ij}$ is a simple part of mixed type, and $C_{ij}(Z)$ is of form $(c\cdot\delta_{mn})$, c an element of K, if $\mathfrak{C}_{ij}$ is of unmixed type.*

PROOF: Since an element Z of the center must commute with the elements $E_\kappa(mn)$ of $\mathfrak{A}$ ($\kappa = 1, 2, \cdots, k$; $m, n = 1, 2, \cdots, f_\kappa$) an easy calculation shows that in terms of the Cartan basis elements

$$Z = \sum_{r,\tau} c^\tau(Z)E_\kappa(r1)B^\tau_{\kappa\kappa}E_\kappa(1r),$$

[11] If we do not take the constituents of $\mathfrak{A}$ identical with constituents of $\mathfrak{A}$ then it is necessary to apply suitable left and right non-singular matrix multipliers to the $\mathfrak{H}^\mu_{\kappa\lambda}$ in the expression for $\bar{\mathfrak{H}}$.

in which the range for τ is determined by the s basis elements B^τ of unmixed type. It follows that $H^\mu(Z) = 0$ if $\mathfrak{H}^\mu$ is of mixed type, and

$$H^\mu(Z) = (c^\mu(Z)\delta_{ij})_{ij}$$

if $\mathfrak{H}^\mu$ is of unmixed type. Applying Theorem 4 we obtain the theorem.

It should be noted that this theorem can be obtained directly from Schur's Lemma and (2).

6

If an element is of type (κ, λ) we say that κ, λ are the type indices of the element. We say that two non-zero simple parts $\mathfrak{C}_{ij}$, $\mathfrak{C}_{pq}$, belong to the same block[12] if there exists a chain of non-zero parts such that any two neighboring parts have at least one type index in common. This relation is reflexive, symmetric, and transitive so that a separation of the non-zero simple parts into distinct disjoint classes is obtained.

THEOREM 10: *Each simple part in the i^{th} row of $\mathfrak{A}$ is either a 0-part or belongs to the same block as $\mathfrak{C}_{ii}$. Similarly, each simple part in the j^{th} column is either a 0-part or belongs to the same block as $\mathfrak{C}_{jj}$. If $\mathfrak{C}_{ii}$, $\mathfrak{C}_{jj}$ do not belong to the same block, then $\mathfrak{C}_{ij}$ is a 0-part.*

For if the simple part $\mathfrak{C}_{ij}$ is not a 0-part, then $\mathfrak{C}_{ii}$, $\mathfrak{C}_{ij}$, $\mathfrak{C}_{jj}$ is a proper chain and so $\mathfrak{C}_{ii}$, $\mathfrak{C}_{ij}$, $\mathfrak{C}_{jj}$ all belong to the same block.

In the same way we may classify the basis elements

$$B^\mu_{\kappa\lambda}, \qquad (\mu = 1, 2, \cdots, m;\ \kappa, \lambda = 1, 2, \cdots, \text{or } k) \tag{11}$$

into blocks. We say that $B^\mu_{\kappa\lambda}$, $B^\gamma_{\rho\sigma}$ belong to the same block if there exists a chain of elements $B^\tau_{\alpha\beta}$ of the set (11) connecting $B^\mu_{\kappa\lambda}$, $B^\gamma_{\rho\sigma}$ such that any two neighboring elements in this chain have at least one type index in common. Let us denote by $\mathfrak{B}$ the set consisting of all the $B^\tau_{\rho\sigma}$ belonging to a given block and their associated elements of form $E_\rho(a1)B^\tau_{\rho\sigma}E_\sigma(1b)$. If A is any element of $\mathfrak{A}$, then $A \cdot E_\rho(a1)B^\tau_{\rho\sigma}E_\sigma(1b)$ is expressible in terms of $\mathfrak{B}$ again; the similar statement is true for the product $E_\rho(a1)B^\tau_{\rho\sigma}E_\sigma(1b) \cdot A$. It follows that the elements of $\mathfrak{B}$ form a basis of an invariant subalgebra of $\mathfrak{A}$: we denote this invariant subalgebra by $\mathfrak{A}^+$.

We can now prove

THEOREM 11: *By elementary transformations $\mathfrak{A}$ may be decomposed into the form*

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{B}_1 & & & 0 \\ & \mathfrak{B}_2 & & \\ & & \ddots & \\ 0 & & & \mathfrak{B}_l \end{pmatrix} \tag{12}$$

where each $\mathfrak{B}_r$ contains all simple parts belonging to one block. The elements of $\mathfrak{A}$ which have only zeros in the parts $\mathfrak{B}_j$, $j \neq i$, form an invariant subalgebra $\mathfrak{A}_i$ of $\mathfrak{A}$,

[12] R. Brauer and C. Nesbitt, *On the Modular Representations of Finite Groups*, Univ. of Toronto Studies, Math. Series, No. 4 (1937); R. Brauer and C. Nesbitt, *On the Modular Character of Groups*, these Annals, Vol. 42 (1941), pp. 556–590; T. Nakayama, *Some Studies on Regular Representations, Induced Representations, and Modular Representations*, these Annals, Vol. 39 (1938), Theorem 5.

and $\mathfrak{A}$ is the direct sum of these invariant subalgebras $\mathfrak{A}_i$. The $\mathfrak{A}_i$ may not be decomposed into a direct sum of invariant subalgebras.

PROOF: If $\mathfrak{C}_{ii}$, $\mathfrak{C}_{i+1,i+1}$ do not belong to the same block, then by Theorem 10, $\mathfrak{C}_{i+1,1}$ is a 0-part. It follows that by permuting the i^{th} row with the $(i+1)^{\text{st}}$ row and the i^{th} column with the $(i+1)^{\text{st}}$ column that we may reverse the positions of $\mathfrak{C}_{ii}$, $\mathfrak{C}_{i+1,i+1}$ in the main diagonal of $\mathfrak{A}$, but leave $\mathfrak{A}$ in reduced form. By successive transformations of this form we may bring together in an upper part $\mathfrak{B}_1$ of $\mathfrak{A}$ all $\mathfrak{C}_{ii}$ which belong to the same block as $\mathfrak{C}_{11}$. We then have

$$\mathfrak{A} = \begin{pmatrix} \mathfrak{B}_1 & 0 \\ \mathfrak{D}_{21} & \mathfrak{D}_{22} \end{pmatrix}.$$

But since the simple parts $\mathfrak{C}_{jj}$ in $\mathfrak{B}_1$ and the $\mathfrak{C}_{ii}$ in $\mathfrak{D}_{22}$ belong now to different blocks, then the simple parts in $\mathfrak{D}_{21}$ must all be 0-parts, or, that is, $\mathfrak{D}_{21}$ is a 0-part. By continuing this process we obtain $\mathfrak{A}$ in the form (12).

To prove the remaining statements in the theorem we first remark that to each basis element $B^{\mu}_{\kappa\lambda}$ there corresponds at least one non-zero simple part $\mathfrak{C}_{ij}$ of type (κ, λ). For, by definition, $B^{\mu}_{\kappa\lambda}$ is a matrix of $\mathfrak{A}$ having non-zero coefficients only in the upper left corners of simple parts of type (κ, λ).

Let now $B^{\tau}_{\rho\sigma}$ be an element of the set $\mathfrak{B}$ of Cartan basis elements which belong to a given block. If the simple parts of type (ρ, σ) are in part $\mathfrak{B}_i$ of $\mathfrak{A}$, then by the above remark $B^{\tau}_{\rho\sigma}$ is an element of the invariant subalgebra $\mathfrak{A}_i$. If $B^{\gamma}_{\mu\nu}$ is also from the set $\mathfrak{B}$, that is, if $B^{\gamma}_{\mu\nu}$ belongs to the same block as $B^{\tau}_{\rho\sigma}$, then the simple parts of type (μ, ν) belong to the same block as simple parts of type (ρ, σ). For a simple part of type (ρ, σ) may be connected by a chain of non-zero simple parts to a simple part of type (μ, ν), the members of this chain corresponding to the members of the chain connecting $B^{\tau}_{\rho\sigma}$, $B^{\gamma}_{\mu\nu}$. Then the simple parts of type (μ, ν) are in the part $\mathfrak{B}_i$ and so $B^{\gamma}_{\mu\nu}$ also is an element of $\mathfrak{A}_i$. It follows that the invariant subalgebra $\mathfrak{A}^{+}$, generated by the elements of $\mathfrak{B}$, coincides with $\mathfrak{A}_i$. We now have at our disposal a correspondence between invariant subalgebras determined by blocks of Cartan basis elements on the one hand, and by blocks of simple parts on the other.

Now let

$$\mathfrak{A} = \bar{\mathfrak{A}}_1 + \bar{\mathfrak{A}}_2 + \cdots + \bar{\mathfrak{A}}_q \tag{13}$$

be a direct decomposition of $\mathfrak{A}$ into invariant subalgebras which may not be directly decomposed, and let $\bar{A}_j$ be an element of $\bar{\mathfrak{A}}_j$. If in the expression

$$\bar{A}_j = \sum h^{\mu}_{ab}(\bar{A}_j) E_{\kappa}(a1) B^{\mu}_{\kappa\lambda} E_{\lambda}(1b)$$

for $\bar{A}_j$ in terms of the Cartan basis elements, $h^{\mu}_{ab}(\bar{A}_j) \neq 0$, then since

$$E_{\kappa}(1a)\bar{A}_j E_{\lambda}(b1) = h^{\mu}_{ab}(\bar{A}_j) B^{\mu}_{\kappa\lambda}$$

is in $\bar{\mathfrak{A}}_j$, so also is $B^{\mu}_{\kappa\lambda}$. In this way we can determine the direct summand of the decomposition (13) to which each element $B^{\mu}_{\kappa\lambda}$ belongs. In particular, we may determine the summands to which the elements $E_{\kappa}(11)$, $(\kappa = 1, 2, \cdots, k)$ belong. If $B^{\mu}_{\kappa\lambda}$ belongs to $\bar{\mathfrak{A}}_j$ then $E_{\kappa}(11)$, $E_{\lambda}(11)$ must also belong to $\bar{\mathfrak{A}}_j$, since

$$E_{\kappa}(11) B^{\mu}_{\kappa\lambda} = B^{\mu}_{\kappa\lambda} E_{\lambda}(11) = B^{\mu}_{\kappa\lambda}$$

and if, for instance, $E_\kappa(11)$ did not belong to the same summand as $B^\mu_{\kappa\lambda}$, the product $E_\kappa(11)B^\mu_{\kappa\lambda}$ would be zero.

Suppose now that the element $B^\tau_{\rho\sigma}$ of the set $\mathfrak{B}$ is in $\bar{\mathfrak{A}}_j$ but that there are elements of $\mathfrak{B}$ which are not in $\bar{\mathfrak{A}}_j$. Then we may choose an element $B^\gamma_{\mu\nu}$ of $\mathfrak{B}$ which is not in $\bar{\mathfrak{A}}_j$, but such that in the chain joining $B^\tau_{\rho\sigma}$ to $B^\gamma_{\mu\nu}$, $B^\gamma_{\mu\nu}$ is the only member which does not belong to $\bar{\mathfrak{A}}_j$. The member preceding $B^\gamma_{\mu\nu}$ in this chain has either the type index μ or ν; let this member be $B^\delta_{\mu\lambda}$. Then by the above, $E_\mu(11)$ is in $\bar{\mathfrak{A}}_j$, since $B^\delta_{\mu\lambda}$ is in $\bar{\mathfrak{A}}_j$; on the other hand $E_\mu(11)$ must be in the same summand as $B^\gamma_{\mu\nu}$, which gives a contradiction. Thus all elements of the set $\mathfrak{B}$ are in $\bar{\mathfrak{A}}_j$. Similarly, if any member of a second set $\mathfrak{B}_1$ of Cartan basis elements which all belong to one block is in $\bar{\mathfrak{A}}_j$, then all members of the set $\mathfrak{B}_1$ are in $\bar{\mathfrak{A}}_j$. But then the invariant subalgebras $\mathfrak{A}^+$, $\mathfrak{A}_1^+$ which have as bases the sets $\mathfrak{B}$, $\mathfrak{B}_1$, respectively, would be direct summands of $\bar{\mathfrak{A}}_j$ contrary to our assumption that $\bar{\mathfrak{A}}_j$ is directly indecomposable. We now have $\mathfrak{A}_i = \mathfrak{A}^+ = \bar{\mathfrak{A}}_j$ which completes our proof.

We suppose now that each $\mathfrak{B}_i$ has been split into its indecomposable parts[13] and that these are in reduced form. Let $U^i_1, U^i_2, \cdots, U^i_n$ be the indecomposable parts of $\mathfrak{B}_i$. We now classify the U^i by the following relation: we say that U^i_ρ, U^i_σ belong to the same block if there exists a chain[14]

$$U^i_\rho, U^i_\alpha, \cdots, U^i_\beta, U^i_\sigma \tag{14}$$

such that any two neighboring parts of the chain (14) have at least one irreducible constituent in common. We will show that all indecomposable parts U^i_ρ ($\rho = 1, 2, \cdots, n$) of $\mathfrak{B}_i$ belong to one block. Let us take together all U^i_ρ that are connected with U^i_1 by such a chain and suppose U^i_α does not belong to this set. Then the simple parts of U^i_α cannot be connected by a proper chain of non-zero simple parts of U^i_1, contrary to our provision that simple parts of $\mathfrak{B}_i$ all belong to one block.

Since a matrix commuting with an indecomposable part U^i_ρ can have just one distinct characteristic root[15] and since, as we have seen, all U^i_ρ of $\mathfrak{B}_i$ have at least one irreducible constituent in common, then for an element Z of the center of $\mathfrak{A}$ the matrix $B_i(Z)$ of $\mathfrak{B}_i$ corresponding to Z has just one distinct characteristic root.

We have then proved

THEOREM 12: *The indecomposable constituent of a part $\mathfrak{B}_i$ may be characterized as belonging to a block. It follows that for a center element Z, the corresponding matrix $B_i(Z)$ has only one distinct characteristic root.*

UNIVERSITY OF ALABAMA

[13] For definitions of indecomposable parts, see B.N., or N.R.

[14] R. Brauer and C. Nesbitt, *On the Modular Representations of Groups of Finite Order*, Univ. of Toronto Studies, Math. Series, No. 4 (1937), p. 14.

[15] R. Brauer and I. Schur, *Zum Irreduzibilitätsbegriff in der Theorie der Gruppen linearer homogener Substitutionen*, Sitzber. der Preuss. Akad. der Wiss., Phys-Math. Klasse, Berlin, XIV (1930).

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

QUADRATIC FORMS PERMITTING COMPOSITION[1]

BY A. A. ALBERT

(Received November 8, 1941)

1. Introduction

An associative algebra, with a unity quantity, over a field $\mathfrak{F}$ has an *involution*[2] (involutorial anti-automorphism) J such that $x + x^J$ and xx^J are in $\mathfrak{F}$ if it is either $\mathfrak{F}$, an algebra of order two over $\mathfrak{F}$, a purely inseparable field the squares of whose elements are in $\mathfrak{F}$ of characteristic two, or a generalized quaternion algebra over $\mathfrak{F}$. The quaternion algebras have order four and may be imbedded in a generalized Cayley-Dickson[3] algebra of order eight over $\mathfrak{F}$. All the algebras so obtained are then *alternative*[4] algebras with a unity quantity and an involution J such that $x + x^J$ and xx^J are in $\mathfrak{F}$. Then the *norm* xx^J is a quadratic form in the coordinates of x which permits composition.

In 1898 A. Hurwitz[5] showed that if $\mathfrak{F}$ is the field $\mathfrak{C}$ of all complex numbers, forms equivalent to the quadratic norm forms described above are the only quadratic forms permitting composition. L. E. Dickson[6] pointed out the connection of these forms with the corresponding algebras and thus completed the proof in the case $\mathfrak{F} = \mathfrak{C}$ of the following

THEOREM. *A quadratic form over $\mathfrak{F}$ permits composition if and only if it is equivalent in $\mathfrak{F}$ to the norm form xx^J of an alternative algebra over $\mathfrak{F}$ with a unity quantity and an involution J such that $x + x^J$ and xx^J are in $\mathfrak{F}$. These latter* **norm** *forms are quadratic forms in* 1, 2, 4 *or* 8 *indeterminates except for the diagonal norm forms, in 2^t indeterminates, of purely inseparable fields of degree 2^t and exponent two over $\mathfrak{F}$ of characteristic two.*

It is readily verified that Dickson's version of the Hurwitz proof is valid for any algebraically closed field of characteristic not two. However, there seems to be no treatment of the case $\mathfrak{F}$ of characteristic two, a case presenting interesting new features. We shall derive the results for this case here and shall indeed

[1] Presented to the Society Nov. 21, 1941.

[2] An involution over $\mathfrak{F}$ is a linear transformation J over $\mathfrak{F}$ of an algebra $\mathfrak{A}$ such that $(ab)^J = b^J a^J$ and J^2 is the identity transformation.

[3] For the definition and elementary properties of such algebras see L. E. Dickson, *On quaternions and their generalizations and the history of the eight square theorem*, these Annals, vol. 20 (1919), pp. 155–71.

[4] See M. Zorn, *Theorie der Alternativen Ringe*, Hamburg Abh. vol. 8 (1930), pp. 123–47 and his *Alternativkörper und Quadratische Systeme*, loc. cit. vol. 9 (1933), pp. 395–402 for a bibliography and the discussion of such algebras.

[5] Göttingen Nachrichten, 1898, pp. 309–16.

[6] See footnote 3.

provide an elegant unified study[7] for the case where $\mathfrak{F}$ is arbitrary. Our results will include a generalization[8] of the Cayley-Dickson algebras to provide new algebras of order 2^e over $\mathfrak{F}$. Certain of these algebras $\mathfrak{B}$ have connected quadratic norm forms and the property that if these norm forms are not null forms then every non-scalar quantity of $\mathfrak{B}$ defines a quadratic subfield. However $\mathfrak{B}$ need not then be a division algebra[9] unless $e < 4$.

2. The general Hurwitz problem

Let $x_1, \cdots, x_n, y_1, \cdots, y_m$ be independent indeterminates over a field $\mathfrak{F}$, $\mathfrak{H} = \mathfrak{F}(x_1, \cdots, x_n, y_1, \cdots, y_m)$ be the corresponding rational function field. Designate the transpose of any matrix S with elements in $\mathfrak{H}$ by S' and define the one rowed matrices

$$x = (x_1, \cdots, x_n), \qquad y = (y_1, \cdots, y_m).$$

Then if A is any n-rowed square matrix with elements in $\mathfrak{F}$ the matrix product $f(x) = xAx'$ is a quadratic form in $x_1, \cdots, x_n$. Conversely every quadratic form $f(x)$ is expressible as such a product. However A is not unique and $xAx' = xA_0x'$ if and only if $A - A_0$ is an alternate (that is, skew-symmetric) matrix.

Suppose that B is an m-rowed square matrix so that $g(y) = yBy'$ is a second quadratic form. Then the general Hurwitz problem[10] is that of determining under what conditions on $f(x)$ and $g(y)$ there exist quantities in $\mathfrak{F}$ such that

$$f(x)g(y) = f(z), \tag{1}$$

where $z = (z_1, \cdots, z_n)$ and

$$z_k = \sum_{\substack{i=1,\cdots,n \\ j=1,\cdots,m}} x_i \gamma_{ik}^{(j)} y_j \qquad (k = 1, \cdots, n). \tag{2}$$

Write $G_j = (\gamma_{ik}^{(j)})$, so that G_j is an n-rowed square matrix with elements in $\mathfrak{F}$. Define

---

[7] The argument required to make the extension of our theorem from the case where our field of reference is an algebraically closed field of characteristic not two to that where it is an arbitrary field seems not to be in the literature and will be a major contribution of our discussion.

[8] The formulation by L. E. Dickson of the Cayley algebras is sufficiently general so that we are able to use it to define such algebras over a field of characteristic two. These algebras are new, however, and so are the other more general types (for example, of order 16) yielded by the Dickson process and given here.

[9] We define algebras which are generalizations of the Cayley-Dickson algebras to algebras of order 2^t and with an involution J such that $x + x^J$ and $xx^J = x^Jx$ are in $\mathfrak{F}$. For proper choice of the defining parameters xx^J is a quadratic form in 2^t squares. If $\mathfrak{F}$ is the field of all real numbers each element of $\mathfrak{B}$ defines a subalgebra equivalent to the complex field. But $\mathfrak{B}$ is not a division algebra for $t > 3$. We shall show this in a later paper where the conditions that our algebras be division algebras will be discussed.

[10] See his posthumous paper in the Math. Annalen, vol. 88 (1922), pp. 1–25.

$$G_y = G_1 y_1 + \cdots + G_m y_m , \tag{3}$$

so that if $e_i = (0, 0, \cdots, 1, 0, \cdots, 0)$ with 1 in the i^{th} column then

$$G_{e_i} = G_i . \tag{4}$$

We see that (2) is equivalent to the statement that z is the matrix product

$$z = xG_y . \tag{5}$$

Then (1) holds if and only if $x[g(y)A]x' = xG_yA(G_y)'x'$, that is, if and only if

$$N = g(y)A - G_yA(G_y)' \tag{6}$$

is an alternate matrix. Thus (2) holds if and only if

$$g(y)E = G_yE(G_y)', \qquad E = A + A', \tag{7}$$

and the diagonal elements of N are all zero.

Let $\phi(\xi) = \xi C\xi'$ and $\psi(\zeta) = \zeta D\zeta'$ be equivalent to $f(x)$ so that there exist nonsingular matrices P and Q such that $\xi C\xi' = \xi PAP'\xi'$, $\zeta D\zeta' = \zeta QAQ'\zeta'$. Then $N_1 = C - PAP'$ and $N_2 = D - QAQ'$ are alternate matrices. If (1) and (2) hold the matrix N of (6) is alternate and so is $PNP' = g(y)PAP' - \Gamma_yQAQ'(\Gamma_y)'$, where $\Gamma_y = PG_yQ^{-1}$. Then the matrix

$$N_0 = g(y)C - \Gamma_yD\Gamma_y' = PNP' + g(y)N_1 - \Gamma_yN_2(\Gamma_y)'$$

is also alternate. Let also $\rho(\eta)$ be equivalent to $g(y)$ so that there exists a nonsingular matrix S such that $g(\eta S) = \rho(\eta)$. Then if $\Delta_\eta = \Gamma_y$, $y = \eta S$, the matrix

$$\rho(\eta)C - \Delta_\eta D(\Delta_\eta)' = N_0$$

is alternate and

$$\phi(\xi)\rho(\eta) = \psi(\zeta) \tag{8}$$

for $\zeta = (\zeta_1, \cdots, \zeta_n)$ and the ζ_k bilinear forms with coefficients in $\mathfrak{F}$ in the ξ_i and the η_j. Thus we have shown that (1) is possible if and only if (8) is possible for quadratic forms $\phi(x)$ and $\psi(x)$ equivalent to $f(x)$ and $\rho(\eta)$ equivalent to $g(y)$.

The principal purpose of our discussion is the derivation of a complete solution of the Hurwitz problem in the case $m = n$ and $g(x)$ equivalent to $f(x)$. Thus we shall stress the study of

$$f(x)f(y) = f(z), \tag{9}$$

for z as in (2) with $m = n$. If $f(x)$ has the property (9) we shall say that *the quadratic form $f(x)$ permits composition.* Then our principal object will be that the determination of the conditions on $f(x)$ that it permit composition.

3. The non-singular case

Let $f(x)$ be a quadratic form given by

$$f(x) = \sum_{i,k=1}^{n} x_i \alpha_{ik} x_k \qquad (\alpha_{ik} \text{ in } \mathfrak{F}), \tag{10}$$

so that $f(x) = xAx'$ for A the n-rowed square matrix (α_{ik}). If the characteristic of $\mathfrak{F}$ is not two we may take A to be a non-singular diagonal matrix so that $a_{ik} = 0$ for $i \neq k$, $a_{ii} \neq 0$. However, when $\mathfrak{F}$ has characteristic two the rank of $A + A'$ is an even integer $2r$, and $n \geqq 2r$. Then[11]

$$f(x) = \alpha_1 x_1^2 + \cdots + \alpha_n x_n^2 \qquad (\alpha_i \neq \text{ in } \mathfrak{F}) \tag{11}$$

if $r = 0$, and otherwise

$$f(x) = \alpha_1 x_1^2 + \cdots + \alpha_n x_n^2 + (x_1 x_{r+1} + \cdots + x_r x_{2r}). \tag{12}$$

Here we assume that if $n > 2r$ the quantities $\alpha_{2r+1}, \cdots, \alpha_n$ are *all* not zero. Moreover if $0 < i < r$ and both α_i and α_{i+r} are zero the transformation $x_i = \xi_i$, $x_{i+r} = \xi_i + \xi_{i+r}$ gives $x_i x_{i+r} = \xi_i^2 + \xi_i \xi_{i+r}$. If $\alpha_i = 0$, $\alpha_{i+r} \neq 0$ we may interchange the corresponding indeterminates. Thus we may assume in all cases that *if $r > 0$ the quantities $\alpha_1, \cdots, \alpha_r$ are all not zero.*

If $\mathfrak{F}$ has characteristic not two and A is a non-singular diagonal matrix as above then $E = A + A' = 2A$ is non-singular. However when $\mathfrak{F}$ has characteristic two the matrix E is non-singular if and only if $n = 2r$,

$$A = \begin{pmatrix} D_1 & I_r \\ 0 & D_2 \end{pmatrix}, \qquad E = A + A' = \begin{pmatrix} 0 & I_r \\ I_r & 0 \end{pmatrix}, \tag{13}$$

where D_1 and D_2 are diagonal matrices and D_1 is non-singular. We shall show later that a quadratic form $f(x)$ permits composition only if it is either a diagonal form (11) or is equivalent to a form xAx' with $A + A'$ non-singular. Let us then study (1) temporarily only in this *non-singular case*, and treat the remaining cases later.

If we replace $y = (y_1, \cdots, y_m)$ in $g(y)$ by $\eta = (\eta_1, \cdots, \eta_m)$ for the η_i in $\mathfrak{F}$ we obtain a quantity $\gamma = g(\eta)$ in $\mathfrak{F}$ which we say is *represented by $g(y)$*. Then we have

LEMMA 1. *Let $\gamma \neq 0$ in $\mathfrak{F}$ be any quantity represented by $g(y)$ in the non-singular case of* (1). *Then $f(x)$ is equivalent to $\gamma^{-1} f(x)$.*

For by (1) we have $f(x)\gamma = f(u)$ where $u = xP$ and $P = G_\eta$. Thus $f(x) = \gamma^{-1} f(u)$ is equivalent to $\gamma^{-1} f(x)$ if we can show that the matrix P is non-singular. But $x\gamma Ax' = xPAP'x'$ if and only if $\gamma A - PAP'$ is alternate. Hence $\gamma E = PEP'$ and P must be non-singular if $\gamma \neq 0$ and E is non-singular.

---

[11] For these results see Chapter II of my *Symmetric and alternate matrices in an arbitrary field*, Transactions of the A. M. S., vol. 43 (1938), pp. 386–436.

Write

$$g(y) = \sum_{i,j=1}^{m} y_i \beta_{ij} y_j , \tag{14}$$

where we are taking $\beta_{11} \neq 0$. Then (1) implies that $\phi(x)g(y) = \phi(z)$ where $\phi(x) = \alpha_{11}^{-1} f(x)$. Also $g(y) = \beta_{11}\psi(y)$ and $\beta_{11} = g(\eta)$, $\eta = (1, 0, \cdots, 0)$. It follows that $\phi(x)$ is equivalent to $\beta_{11}^{-1}\phi(x)$ and that then $[\beta_{11}^{-1}\phi(x)] \cdot [\beta_{11}\psi(y)] = \phi(z)$, $\phi(x)\psi(y) = \phi(z)$. Thus we have shown that in the non-singular case there is no loss of generality if we assume that $\alpha_{11} = \beta_{11} = 1$.

Note that in the study of quadratic forms $f(x)$ permitting composition Lemma 1 implies that $f(x)$ is equivalent to $\alpha_{11}^{-1} f(x)$ and thus that we need only study forms $f(x)$ with $\alpha_{11} = 1$ in the non-singular case.

We now write (7) in the equivalent form

$$G_y (G_y)^J = g(y) I, \tag{15}$$

where I is the n-rowed identity matrix and

$$(G_y)^J = E G_y E^{-1} = G_1^J y_1 + \cdots + G_m^J y_m . \tag{16}$$

It is well known that the correspondence $G \to G^J = EG'E^{-1}$ is an involution of the algebra of all n-rowed square matrices with elements in $\mathfrak{H}$, and we shall use the consequent properties of J.

The equation (7) is an identity in the indeterminates $y_1, \cdots, y_m$ and is equivalent, in view of (14), to

$$G_i G_i^J = \beta_{ii} I,\ G_i G_j^J + G_j G_i^J = \delta_{ij} I \qquad (i \neq j;\ i, j = 1, \cdots, m), \tag{17}$$

where

$$\delta_{ij} = \delta_{ji} = \beta_{ij} + \beta_{ji} . \tag{18}$$

In particular $G_1 G_1^J = I$ since we have taken $\beta_{11} = 1$. Then also $G_1^J G_1 = I$. We now define

$$G_y^{(0)} = G_1^J G_y = G_1^{(0)} y_1 + \cdots + G_m^{(0)} y_m \tag{19}$$

and clearly have

$$G_1^{(0)} = I,\ G_y^{(0)} (G_y^{(0)})^J = G_1^J (G_y G_y^J) G_1 = g(y) G_1^J G_1 = g(y) I, \tag{20}$$

so that $G_y^{(0)}$ is a solution of (7) if G_y is. But also $G_1^J = G_y^{-1}$ and $g(e_1) = 1$, (6) states that $A - G_1 A G_1'$ is alternate and so is $G_1^J A (G_1^J)' - A$. Then by (6)

$$N_0 = [A - G_1^J A (G_1^J)'] g(y) + G_1^J [g(y) A - G_y A (G_y)'] (G_1^J)' = A g(y) - G_y^{(0)} A (G_y^{(0)})'$$

is an alternate matrix. It follows that G_y defines a solution (5) of (2) so that (1) holds if and only if $G_y^{(0)}$ does, and we may thus assume that

$$G_1 = G_1^J = I. \tag{21}$$

Equations (17) for $i = 1$ now give

$$G_j^J = \delta_{1j} I - G_j \qquad (j = 2, \cdots, m), \tag{22}$$

while $G_1^J = 2I - G_1$. But then

$$(G_y)^J = t(G_y) - G_y, \tag{23}$$

where the trace function, $t(G_y)$, is the scalar matrix

$$G_y + (G_y)^J = \left(2y_1 + \sum_{j=2}^{m} \delta_{1j} y_j\right) I, \tag{24}$$

and is linear in the coordinates y_i of y. Thus the mapping

$$\eta \to G_\eta$$

is a linear mapping of the linear space $\mathfrak{L}$ of one by m vectors η on the linear space of matrices G_η. Moreover if we define $e = (1, 0, \cdots, 0)$,

$$\eta^J = \left(2\eta_1 + \sum_{j=2}^{m} \delta_{1j} \eta_j\right) e - \eta \tag{25}$$

we see that J is a linear transformation on $\mathfrak{L}$ such that

$$G_{\eta J} = (G_\eta)^J. \tag{26}$$

But $[(G_\eta)^J]^J = G_\eta$ and hence

$$(\eta^J)^J = \eta. \tag{27}$$

We shall now specialize our study of (1) to the case $m = n$ and $g(y) = f(y)$.

4. Composition in the non-singular case

Let $f(x)$ be a quadratic form permitting composition. Then

$$f(x)f(y) = f(xG_y), \tag{28}$$

where the elements of the n-rowed square matrix G_y are linear forms in $y_1, \cdots, y_n$ and we have seen that if $e = (1, 0, \cdots, 0)$ we may take

$$G_e = I, \qquad f(e) = 1, \qquad G_y G_y^J = f(y) I \tag{29}$$

such that (23) and (24) hold. Then the ordinary matrix product $z = xG_y$ gives (2) and these equations may also be written as the first equation in

$$z = yH_x = xG_y, \tag{30}$$

where we define $\rho_{jk}^{(i)} = \gamma_{ik}^{(j)}$ and have

$$H_i = (\rho_{jk}^{(i)}), \qquad H_x = H_1 x_1 + \cdots + H_n x_n. \tag{31}$$

But $f(y)f(x) = f(yH_x)$ and our derivation of the properties of G_y implies that we may interchange the roles of $f(x)$ and $f(y)$ to obtain

$$H_x H_x^J = f(x) I, \tag{32}$$

as a consequence of (15). It follows from (29) that

$$H_e H_e^J = I, \tag{33}$$

and thus that H_e is non-singular. Define

$$u = u(y) = yH_e, \qquad R_u = G_y, \tag{34}$$

so that $u = (u_1, \cdots, u_n)$, $y = uH_e^{-1}$ and $y_1, \cdots, y_n$ are linear forms in $u_1, \cdots, u_n$. Then by (30) and (28) with $x = e$ we have

$$u = eG_y = eR_u, \qquad f(u) = f(eG_y) = f(e)f(y) = f(y). \tag{35}$$

Moreover $u(e) = eG_e = e$ since $G_e = I$, $R_{u(e)} = G_e = I$. But then $f(x)f(y) = f(xG_y) = f(xR_u) = f(x)f(u)$,

$$f(x)f(y) = f(xR_y). \tag{36}$$

We have thus proved that R_y is a solution G_y of (2) such that

$$R_e = I, \quad eR_y = y, \quad R_y R_y^J = f(y)I, \quad N = f(y)A - R_y A R_y' \tag{37}$$

is an alternate matrix.

Let now $\mathfrak{x}$ be the linear space of all $\xi = (\xi_1, \cdots, \xi_n)$ for ξ_i in $\mathfrak{F}$. Then the correspondence

$$\xi \to R_\xi \tag{38}$$

is evidently a linear mapping of $\mathfrak{x}$ on the linear space of all the matrices R_ξ. Define an operation

$$\xi \cdot \eta = \xi R_\eta \tag{39}$$

on $\mathfrak{x}\mathfrak{x}$ to $\mathfrak{x}$ and call the result $\xi \cdot \eta$ the *product* of ξ and η. Then $\mathfrak{x}$ becomes an algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$. Indeed all algebras may be defined in this way.

We have $e \cdot \eta = eR_\eta = \eta$ by (37), and $\eta \cdot e = \eta R_e = \eta I = \eta$. Hence $\mathfrak{A}$ has e as its unity quantity. We shall give further properties of $\mathfrak{A}$ in the next section. We now relate this result to a certain class of algebras.

An algebra $\mathfrak{A}$ over $\mathfrak{F}$ is called *alternative* if

$$x(xy) = (xx)y, \qquad (yx)x = y(xx) \tag{40}$$

for every x and y of $\mathfrak{A}$. Then $(x + y)[(x + y)x] = (x + y)(xx + yx) = x(xx) + x(yx) + y(xx) + y(yx) = [(x + y)(x + y)]x = (xx + xy + yx + yy)x = (xx)x + (xy)x + (yx)x + (yy)x$. By (40) we have $x(yx) = (xy)x$, a postulate which is sometimes given as one of the alternative postulates but which is actually a consequence of (40). We now prove

LEMMA 2. *Let $\mathfrak{A}$ be an algebra with a unity quantity and an involution J such that*

$$t_x = x + x^J, \qquad xx^J \tag{41}$$

are in $\mathfrak{F}$. Then $x^J x = xx^J$, and $\mathfrak{A}$ is alternative if and only if

$$(xx^J)y = x(x^J y) = x^J(xy) = (yx)x^J = (yx^J)x. \tag{42}$$

For $xx^J = x(t_x - x) = t_x x - xx = (t_x - x)x = x^J x$ since t_x is a scalar. Also

$$x(x^J y) = x[(t_x - x)y] = t_x(xy) - x(xy),$$

and

$$(xx^J)y = [x(t_x - x)]y = t_x(xy) - (xx)y.$$

Thus $x(x^J y) = (xx^J)y$ if and only if $x(xy) = (xx)y$. Similarly $(yx)x^J = y(xx^J)$ if and only if $(yx)x = y(xx)$. However $(x^J)^J = x$ and if $\mathfrak{A}$ is alternative we have $x^J(xy) = (x^J x)y = (xx^J)y$, $(yx^J)x = y(x^J x) = (xx^J)y$.

We next prove

LEMMA 3. *Let $\mathfrak{A}$ be alternative and as in Lemma 1. Then the quadratic form xx^J in the coordinates $x_1, \cdots, x_n$ of the quantities $x = (x_1, \cdots, x_n)$ of $\mathfrak{A}$ permits composition.*

For it is clear that if $\mathfrak{H}$ is any scalar extension of $\mathfrak{F}$ the algebra $\mathfrak{A}_\mathfrak{H}$ has the properties we have assumed for $\mathfrak{A}$. We let $x_1, \cdots, x_n, y_1, \cdots, y_n$ be independent indeterminates over $\mathfrak{F}$ and $\mathfrak{H} = \mathfrak{F}(x_1, \cdots, x_n, y_1, \cdots, y_n)$. Then $x = (x_1, \cdots, x_n)$ and $y = (y_1, \cdots, y_n)$ are non-zero quantities of $\mathfrak{A}_\mathfrak{H}$ and so are x^J, y^J. Also $xx^J = x^J x = f(x)$ is in $\mathfrak{H}$. Then so are $f(y) = yy^J = y^J y, f(xy) = (xy)(xy)^J$. But $[f(xy)]x^J = x^J[(xy)(xy)^J] = [x^J(xy)](y^J x^J) = f(x)[y(y^J x^J)] = f(x)f(y)x^J, f(x)f(y) = f(xy)$ as desired.

We now prove the partial converse,

LEMMA 4. *Let $f(x) = xAx'$ such that $A + A'$ is non-singular. Then $f(x)$ permits composition if and only if $f(x)$ is the norm form xx^J of an alternative algebra $\mathfrak{A}$ with a unity quantity and an involution J such that $x + x^J$ and xx^J are*[12] *in $\mathfrak{F}$.*

For $f(x) = \sum x_i \alpha_{ij} x_j$ with $\alpha_{11} = 1$ and all the α_{ij} in $\mathfrak{F}$ and we define an algebra $\mathfrak{A}$ over $\mathfrak{F}$ by (39) and have (37). Then if $\xi = (\xi_1, \cdots, \xi_n)$ with the ξ_i in $\mathfrak{F}$ we have seen that $R_\xi + (R_\xi)^J = t_\xi I$ where $t = t_\xi$ is in $\mathfrak{F}$ and is determined by $f(x)$, that is, actually

$$t_\xi = 2\xi_1 + \sum_{j=2}^{n} (\alpha_{1j} + \alpha_{j1})\xi_j. \tag{43}$$

We now write

$$\xi^J = t_\xi e - \xi, \tag{44}$$

and have defined a correspondence J which is evidently a linear transformation on $\mathfrak{A}$. Now if $t = t_\xi$ we have

$$R_{\xi^J} = R_{te-\xi} = tI - R_\xi = (R_\xi)^J \tag{45}$$

so that $t_{\xi^J} = t_\xi$, $(\xi^J)^J = \xi$ for every ξ of $\mathfrak{A}$. Since $\xi\eta$ is not defined as a matrix product we may drop the dot in (39). We define $\mathfrak{H} = \mathfrak{F}(x_1, \cdots, x_n,$

[12] In fact we show that each solution G_y of the consequent equation $z = xG_y$ defines an algebra. Moreover those of these algebras which have unity quantities have the property of our lemma.

$y_1, \cdots, y_n$) as above and in $\mathfrak{A}_\mathfrak{S}$ have $x = (x_1, \cdots, x_n)$, $y = (y_1, \cdots, y_n)$, $f(x)e = xx^J$ is the norm form of $\mathfrak{A}$ as desired. By the proof of Lemma 1 $xx^J = x^Jx$. Finally

$$(yx)x^J = (yR_x)(R_{x^J}) = y(R_xR_x^J) = yf(x), \tag{46}$$

so that we have part of (42).

We note now that we may write $\xi\eta = \xi R_\eta = \eta L_\xi$ as in (30). But then $L_\xi + (L_\xi)^J = t_\xi I$ since t_ξ depends only on $\xi_1, \cdots, \xi_n$ and the coefficients of $f(x)$, not on the choice of the solution G_y of (2). Thus $(L_\xi)^J = L_{\xi^J}$. It follows that $x^J(xy) = (yL_x)L_{x^J} = yL_x(L_x)^J = yf(x)$. We now use $f(x)f(y) = f(xy)$ and form $x^Jf(xy) = x^J[(xy)(xy)^J] = x^J(xy)(xy)^J = f(x)y(xy)^J = f(x)f(y)x^J$. Thus $y(xy)^J = (yy^J)x^J = y(y^Jx^J)$. Multiplying by y^J we have $y^Jy(xy)^J = (y^Jy)(y^Jx^J)$, $(xy)^J = y^Jx^J$. But then $(\xi\eta)^J = \eta^J\xi^J$ for every ξ and η of $\mathfrak{A}$, J defines an involution of $\mathfrak{A}$ such that $\xi + \xi^J$ and $\xi\xi^J$ are in $\mathfrak{F}$, $\mathfrak{A}$ is alternative as desired. This completes our proof.

5. Algebras and their norm forms

The proof of our principal theorem requires that we exhibit algebras $\mathfrak{A}$ whose norm forms xx^J are quadratic forms xAx' with $A + A'$ non-singular. We shall show later that such algebras necessarily have orders 1, 2, 4, 8. Let us then exhibit algebras of the kind desired and of these orders.

The algebra of order one is $\mathfrak{F}$ itself. Any algebra $\mathfrak{A} = (1, u)$ with $u^2 = \beta u + \gamma$ and β and γ in $\mathfrak{F}$ has the desired properties and is trivially shown to be an associative algebra. Then if $\beta = 0$ and $\mathfrak{F}$ has characteristic two we have $u^2 = \gamma$, $(u^J)^2 = \gamma$, $u^J = u$ for any involution J. But then $xx^J = xx = x_1^2 + x_2^2\gamma = xAx'$ with $A + A' = 0$, contrary to hypothesis. If $\beta = 0$ and $\mathfrak{F}$ does not have characteristic two we have $(u + 1)^2 = u^2 + 2u + 1 = 2(u + 1) + \gamma - 1$ and may take $\beta \neq 0$. Define $k = \beta^{-1}u$ and have $k^2 = k + \alpha$ for α in $\mathfrak{F}$. Thus the algebra

$$\mathfrak{A} = (1, k), \qquad k^2 = k + \alpha, \tag{47}$$

is the only algebra of order two which we need to consider. It is associative, has the automorphism T defined by

$$x = x_1 + x_2k, \qquad x^T = x_1 + x_2(1 - k), \tag{48}$$

and T is an involution of $\mathfrak{A}$. It follows that

$$x + x^T = 2x_1 + x_2, \qquad xx^T = x_1^2 + x_1x_2 - x_2^2. \tag{49}$$

The *quaternion* algebras may be defined as algebras $\mathfrak{A}$ of matrices

$$x = \begin{pmatrix} a & b \\ b^T\beta & a^T \end{pmatrix}, \tag{50}$$

with $\beta \neq 0$ in $\mathfrak{F}$, $a = x_1 + x_2k$, $b = x_3 + x_4k$, and k and T defined for the algebra

given by (47), (48). Then $\mathfrak{A}$ has a basis 1, u, v, uv over $\mathfrak{F}$, where the unity quantity 1 of $\mathfrak{A}$ is really the two rowed identity matrix, and

$$u = \begin{pmatrix} k & 0 \\ 0 & 1-k \end{pmatrix}, \qquad v = \begin{pmatrix} 0 & 1 \\ \beta & 0 \end{pmatrix}. \tag{51}$$

Then $vu = (1-u)v$, $u^2 = u + 1$, $v^2 = \beta \neq 0$. Conversely every algebra with such a multiplication table is a simple associative algebra of degree two over its center[13] $\mathfrak{F}$, and is representable as the set of matrices (50).

Every involution[14] of a two-rowed total matric algebra is a correspondence

$$x \to x^J = ex'e^{-1}, \tag{52}$$

for a non-singular symmetric or alternate matrix e. We take

$$e = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{53}$$

and see that

$$x^J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} a & b^T\beta \\ b & a^T \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a^T & -b \\ -b^T\beta & a \end{pmatrix}. \tag{54}$$

Then x^J is in $\mathfrak{A}$ and (54) defines an involution over $\mathfrak{F}$ of $\mathfrak{A}$. Moreover if $a = x_1 + x_2k$ we have

$$x + x^J = a + a^T = 2x_1 + x_2 \tag{55}$$

in $\mathfrak{F}$,

$$f(x) = xx^J = \begin{pmatrix} a & b \\ b^T\beta & a^T \end{pmatrix}\begin{pmatrix} a^T & -b \\ -b^T\beta & a \end{pmatrix} = aa^T - \beta bb^T = x^Jx. \tag{56}$$

Thus $f(x) = x_1^2 + x_1x_2 - \alpha x_2^2 - (x_3^2 + x_3x_4 - x_4^2\alpha)\beta$ is the determinant of x and is a multiplicative function, that is, $f(x)f(y) = f(z)$ where $z = xy$.

Forms in $x_1, \cdots, x_8$ have been connected[15] with certain non-associative alternative algebras. We shall construct these algebras over fields of arbitrary characteristic by the use of L. E. Dickson's formulation of the non-associative algebra of Cayley. Let $\mathfrak{A}$ be an algebra of order n over $\mathfrak{F}$ and let $\mathfrak{A}$ have a unity quantity 1 which we shall identify with the unity quantity of $\mathfrak{F}$. Assume that $\mathfrak{A}$ has an involution J over $\mathfrak{F}$ such that $x + x^J$ and $xx^J = x^Jx$ are in $\mathfrak{F}$. Then every x of $\mathfrak{A}$ is a root of an equation

$$\phi(\lambda) = (\lambda - x)(\lambda - x^J) = \lambda^2 - (x + x^J)\lambda + xx^J = 0$$

with coefficients $t(x) = x + x^J$ and $f(x) = xx^J$ in F. Hence $\mathfrak{A}$ has degree two over $\mathfrak{F}$ and $\phi(\lambda)$ is the minimum function of every x of $\mathfrak{A}$ which is not in $\mathfrak{F}$.

[13] We use *center* instead of *centrum*.

[14] See Chapter V of my *Modern Higher Algebra*.

[15] See the paper referred to in footnote 3.

Every x in $\mathfrak{A}$ defines a subalgebra $\mathfrak{F}[x]$ of all polynomials in x and the unity quantity of $\mathfrak{A}$ with coefficients in $\mathfrak{F}$. Clearly if x is not in $\mathfrak{F}$ then $\mathfrak{F}[x] = (1, x)$ over $\mathfrak{F}$ and $\mathfrak{F}[x]$ is an associative subalgebra of $\mathfrak{A}$. Since $x^J = t(x) - x$ we see that the involution J of $\mathfrak{A}$ induces an automorphism in every subalgebra $\mathfrak{F}[x]$ of $\mathfrak{A}$ which is the identity if and only if either x is in $\mathfrak{F}$ or $t(x) = 0$ and $\mathfrak{F}$ has characteristic two. Observe now that the function $f(x) = xx^J$ is a quadratic form in the coordinates of x with respect to any basis of $\mathfrak{A}$ and that every $x \neq 0$ of $\mathfrak{A}$ has an inverse in $\mathfrak{A}$ and actually in $\mathfrak{F}[x]$ if and only if $f(x)$ is not a null form.

We shall now construct an algebra $\mathfrak{B}$ of order $2n$ over $\mathfrak{F}$ with the properties described above by the use of any such algebra $\mathfrak{A}$ of order n over $\mathfrak{F}$. The algebra $\mathfrak{B}$ will have quantities which are pairs of quantities of $\mathfrak{A}$ and it is customary to indicate this by the statement that these quantities are uniquely expressible in the form

$$x = g_1 + g_2 w \tag{57}$$

for g_1 and g_2 in $\mathfrak{A}$. Let $\gamma \neq 0$ be in $\mathfrak{F}$, J be the given involution of $\mathfrak{A}$ and define multiplication[16] in $\mathfrak{B}$ by

$$xy = (g_1 + g_2 w)(h_1 + h_2 w) = (g_1 h_1 + \gamma h_2^J g_2) + (h_2 g_1 + g_2 h_1^J)w. \tag{58}$$

Extend the definition of J so that J becomes a linear transformation of $\mathfrak{B}$ by means of

$$x^J = g_1^J - g_2 w. \tag{59}$$

It is clear that J is now a one-to-one linear transformation of $\mathfrak{B}$ such that J^2 is the identity transformation.

To form the product xx^J we replace h_2 by $-g_2$, h_1^J by g_1 in (58) and have $h_2 g_1 + g_2 h_1^J = -g_2 g_1 + g_2 g_1 = 0$. Then

$$xx^J = g_1 g_1^J - \gamma g_2 g_2^J = x^J x \tag{60}$$

since the interchange of g_1 with g_1^J , g_2 with $-g_2$ does not alter $g_1 g^J - \gamma g_2 g_2^J$. Also

$$x + x^J = g_1 + g_1^J \tag{61}$$

is in $\mathfrak{F}$, each x of $\mathfrak{B}$ is a root of $\lambda^2 - (x + x^J)\lambda + xx^J = 0$, and $\mathfrak{B}$ has degree two over $\mathfrak{F}$. Finally

$$(xy)^J = (h_1^J g_1^J + \gamma g_2^J h_2) - (h_2 g_1 + g_2 h_1^J)w, \tag{62}$$

and

$$\begin{aligned} y^J x^J &= (h_1^J - h_2 w)(g_1^J - g_2 w) \\ &= (h_1^J g_1^J + \gamma g_2^J h_2) - (g_2 h_1^J + h_2 g_1)w, \end{aligned} \tag{63}$$

[16] This is clearly a generalization of the L. E. Dickson formulation of the definition of the Cayley algebra of order eight.

so that $(xy)^J = y^J x^J$ and J is an involution of $\mathfrak{B}$. We observe that (58) may be given more simply by the use of the distributive law and the relations

$$g(hw) = (hg)w, \qquad (gw)h = (gh^J)w,$$
$$(gw)(hw) = \gamma h^J g, \tag{64}$$

for every g and h of $\mathfrak{A}$. Clearly if $\mathfrak{A}$ is not a commutative algebra the algebra $\mathfrak{B}$ is not associative.

To prove that $\mathfrak{B}$ is alternative we use (57) to compute

$$x(xy) = [g_1(g_1h_1 + \gamma h_2^J g_2) + \gamma(g_1^J h_2^J + h_1 g_2^J)g_2] + [(h_2g_1 + g_2h_1^J)g_1 + g_2(h_1^J g_1^J + \gamma g_2^J h_2)]w, \tag{65}$$

as well as $x^2 = g_1^2 + \gamma g_2^J g_2 + [g_2(g_1 + g_1^J)]w$ and

$$x^2y = \{(g_1^2 + \gamma g_2^J g_2)h_1 + \gamma h_2^J[g_2(g_1 + g_1^J)]\} + \{h_2(g_1^2 + \gamma g_2^J g_2) + [g_2(g_1 + g_1^J)]h_1^J\}w. \tag{66}$$

If $\mathfrak{A}$ is an alternative algebra we have $g_1(g_1h_1) + \gamma(h_1g_2^J)g_2 = g_1^2h_1 + \gamma(g_2^J g_2)h_1$. However $g_1 + g_1^J$ is in $\mathfrak{F}$ and $h_2^J[g_2(g_1 + g_1^J)] = (g_1 + g_1^J)(h_2^J g_2) = g_1(h_2^J g_2) + (g_1^J h_2^J)g_2$ if and only if $g_1^J(h_2^J g_2) = (g_1^J h_2^J)g_2$, that is, $\mathfrak{A}$ is an associative algebra. Conversely if $\mathfrak{A}$ is associative we have also $(h_2g_1)g_1 + \gamma g_2(g_2^J h_2) = h_2(g_1^2 + \gamma g_2^J g_2)$, $(g_2h_1^J)g_1 + g_2(h_1^J g_1^J) = g_2(g_1 + g_1^J)h_1^J$, and $x(xy) = x^2y$ for every x and y of $\mathfrak{B}$.

It follows that $x^J(x^J y^J) = (x^J x^J)y^J$ and, by operating with J, that $(yx)x = y(xx)$ for every x and y of $\mathfrak{B}$. Hence $\mathfrak{B}$ is alternative. This completes our proof of the result that $\mathfrak{B}$ is an alternative algebra if and only if $\mathfrak{A}$ is associative.

We now see that if $\mathfrak{A}$ is the associative algebra of order four defined by the matrices of (50) the algebra $\mathfrak{B}$ is an alternative algebra of order eight over $\mathfrak{F}$. Its norm form $f(x) = xx^J = x^J x$ is given explicitly by

$$f(x) = x_1^2 + x_1x_2 - x_2^2\alpha - (x_3^2 + x_3x_4 - x_4^2\alpha)\beta - (x_5^2 + x_5x_6 - x_6^2\alpha)\gamma + (x_7^2 + x_7x_8 - x_8^2\alpha)\beta\gamma. \tag{67}$$

Note that $f(x)$ becomes the norm form of the algebra $\mathfrak{A}$ of (50) by putting $x_5 = x_6 = x_7 = x_8 = 0$, and that of (47) by taking also $x_3 = x_4 = 0$. The values $\alpha = \beta = \gamma = -1$ then give us the forms $x_1^2 + x_2^2 + x_1x_2$, $x_1^2 + \cdots + x_4^2 + x_1x_2 + x_3x_4$, $x_1^2 + \cdots + x_8^2 + x_1x_2 + x_3x_4 + x_5x_6 + x_7x_8$.

Observe that the algebra $\mathfrak{B}$ of order eight defined above is a division algebra if and only if its associative subalgebra $\mathfrak{A}$ is a division algebra and $f(x)$ is not a null form. For if $xy = 0$ then $x^J(xy) = f(x)y = 0$ and if $x \neq 0$ then $f(x) \neq 0$, $y = 0$. The condition that $xx^J = 0$ if and only if $x = 0$ is satisfied when γ is chosen so that γ is not the norm gg^J of any g of $\mathfrak{A}$.

It is already known that in $\mathfrak{F}$ of characteristic not two a quadratic form $f(x) = xAx'$ with $A = A'$ a non-singular n-rowed matrix permits composition only when $n = 1, 2, 4, 8$. Thus it remains to show the corresponding result for

$A + A'$ non-singular and $\mathfrak{F}$ of characteristic two, as well as to study other possible cases of composition for this characteristic. Let us thus assume henceforth that the characteristic of $\mathfrak{F}$ is two.

6. Determination of possible orders in the non-singular case

Let $\mathfrak{A}$ be any simple algebra of degree n over its center $\mathfrak{F}$ and let I be the unity quantity of $\mathfrak{A}$. Assume the existence of quantities U and V in $\mathfrak{A}$ such that

$$U^2 = \beta I, \qquad V^2 = \gamma I, \qquad W = UV = I - VU, \tag{68}$$

for $\beta \neq 0$ and γ in $\mathfrak{F}$. Then $W^2 = U(I - UV)V = W - \beta\gamma I$, $UW + WU = \beta V + U(I - UV) = U$. It follows that

$$W^2 = W - \beta\gamma I, \qquad UW = (I - W)U. \tag{69}$$

But then $\mathfrak{A}$ contains a generalized quaternion algebra $\mathfrak{B} = (I, W, U, WU) = (I, U, V, UV)$. This algebra of order four over $\mathfrak{F}$ is a simple algebra of degree two over its center $\mathfrak{F}$. But then it is well known[17] that $\mathfrak{A}$ is the direct product

$$\mathfrak{A} = \mathfrak{B} \times \mathfrak{A}_1, \qquad \mathfrak{A}_1 = \mathfrak{A}^{\mathfrak{B}}, \tag{70}$$

where we write $\mathfrak{A}^{\mathfrak{B}}$ for the set of all quantities of $\mathfrak{A}$ commutative with both U and V. Moreover $n = 2n_1$, $\mathfrak{A}_1$ is a simple algebra of degree n_1 over its center $\mathfrak{F}$. Assume now that $\mathfrak{A}$ contains also U_2, V_2 such that $U_2^2 = \beta_2 I$, $V_2^2 = \gamma_2 I$, $U_2V_2 + V_2U_2 = I$, where $\beta_2 \neq 0$ and γ_2 are in $\mathfrak{F}$. Then if $UU_2 - U_2U = UV_2 - V_2U = VU_2 - U_2V = VV_2 - V_2V = 0$, the quantities U_2 and V_2 are in $\mathfrak{A}^{\mathfrak{B}}$. But then $\mathfrak{A}_1$ contains $\mathfrak{B}_2 = (I, U_2, V_2, U_2V_2)$, and $\mathfrak{A}_1 = \mathfrak{B}_2 \times \mathfrak{A}_2$ where $n_1 = 2n_2$, $\mathfrak{A}_2 = \mathfrak{A}^{\mathfrak{B}_2}$ is simple of degree n_2 over its center $\mathfrak{F}$. We now let $s \geqq 1$ and suppose that $\mathfrak{A}$ contains $s - 1$ pairs of quantities U_j, V_j such that

$$U_j^2 = \beta_j I, \quad V_j^2 = \gamma_j I, \quad U_jV_j + V_jU_j = I \qquad (j = 1, \cdots, s - 1), \tag{71}$$

for $\beta_j \neq 0$ and γ_j in $\mathfrak{F}$. Assume also that

$$U_jU_k - U_kU_j = U_jV_k - V_kU_j = V_jV_k - V_kV_j = 0 \qquad (j \neq k; j, k = 1, \cdots, s - 1). \tag{72}$$

Then $\mathfrak{A} = \mathfrak{B}_1 \times \cdots \times \mathfrak{B}_{s-1} \times \mathfrak{A}_s$, and we have proved that 2^{s-1} divides n.

Let us now apply these results to the study of (1) in the case where $f(x) = xAx'$, $E = A + A'$ is non-singular, and we have (12) for $n = 2r$, $\alpha_1 = 1$. We assume that $\mathfrak{F}$ has characteristic two and that

$$g(y) = y_1^2 + \beta_2 y_2^2 + \cdots + \beta_m y_m^2 + (y_1y_{s+1} + \cdots + y_sy_{2s}) \tag{73}$$

where $m \geqq 2s$, the β_j are in $\mathfrak{F}$, and $\beta_2, \cdots, \beta_s$ are all not zero. By (17) there exist matrices $G_1 = I, G_2, \cdots, G_m$ such that

$$G_iG_i^J = \beta_i, \qquad G_iG_k^J + G_kG_i^J = \delta_{ik}I \qquad (i \neq k; i, k = 1, \cdots, m), \tag{74}$$

[17] See my *Structure of Algebras*, pp. 52–55.

where now $\delta_{ik} = 0$ unless $0 < i \leqq 2s$, $0 < k \leqq 2s$ and $i - k$ is s or $-s$. In the remaining cases $\delta_{ik} = 1$. Define $U_{j-1} = G_j$, $V_{j-1} = G_{j+s}$ for $j = 2, \cdots, s$. Then by (74) for $i = 1$ we have $G_k^J + G_k = 0$ for $k = 2, \cdots, s$ and thus $G_k^J = G_k$. Hence $U_j^2 = \beta_{j+1}$, $V_j^2 = \beta_{j+1+s}$ for $j = 1, \cdots, s-1$ where $\beta_{j+1} \neq 0$. Also by (74) for $i = j + 1$ and $k = i + s$ we have $U_jV_j + V_jU_j = I$ and thus (71). By these same equations for $k - i \neq s$ and $i = j + 1$ or $j + 1 + s$ respectively we obtain (72). We let $\mathfrak{A}$ be the algebra of all n-rowed square matrices with elements in $\mathfrak{F}$ and it follows as above that 2^{r-1} divides n.

Note that we have not utilized all consequences of (74) and that the conditions omitted as well as other results would have to be used in order to complete the study of (1). However we are using this study of (1) only as a tool in our discussion of quadratic forms permitting composition and so shall leave the more general considerations for later study.

Let us then take $m = 2s = n = 2r$. The results just derived imply that 2^{r-1} divides $2r$. This is not the case for $r = 3$, $r > 4$, it is the case for $r = 1, 2, 4$, and thus $n = 2, 4, 8$. We have shown that in the non-singular case over $\mathfrak{F}$ of characteristic two a form $f(x) = f(x_1, \cdots, x_n)$ permits composition only if $n = 2, 4, 8$. We combine this with the known results[18] for $\mathfrak{F}$ of characteristic not two to see that in the non-singular case the only possible values of n are 1, 2, 4, 8.

7. The singular case with $A + A' \neq 0$

Let $\mathfrak{F}$ be a field of characteristic two and $f(x) = xAx'$ such that $E = A + A' \neq 0$ is singular. Then $f(x)$ has the form (12) for $n > 2r > 0$ and we have seen that we may take $\alpha_1, \cdots, \alpha_r$ and $\alpha_{2r+1}, \cdots, \alpha_n$ all not zero. Thus

$$A = \begin{pmatrix} A_1 & 0 \\ 0 & D \end{pmatrix}, \qquad A_1 = \begin{pmatrix} D_1 & I_r \\ 0 & D_2 \end{pmatrix}, \tag{75}$$

[18] The method we have used does not require that $\mathfrak{F}$ be algebraically closed and is applicable to obtain a much simpler derivation than that in the literature of the analogous results for $\mathfrak{F}$ of characteristic not two. We take $f(x) = \alpha_1x_1^2 + \cdots \alpha_nx_n^2$, $g(y) = \beta_1y_1^2 + \cdots + \beta_my_m^2$ for the α_i and β_j all not zero and in $\mathfrak{F}$. Then we have seen that we may take $\alpha_1 = \beta_1 = 1$ and have $G_iG_i^J = \beta_iI$, $G_iG_j^J + G_jG_i^J = 0$ for $i \neq j$ and $i, j = 1, \cdots, m$. Thus we may take $G_1 = I$ and obtain $G_i^J = -G_i$, $G_i^2 = -\beta_iI$, $G_iG_j = -G_jG_i$ for $i \neq j$ and $i, j = 2, \cdots, m$. We define μ to be the greatest integer in $\frac{1}{3}m$ and may prove that 2^μ divides n. For $\mathfrak{B}_1 = (I, G_2, G_3, G_2G_3)$ is a generalized quaternion subalgebra of the total matric algebra $\mathfrak{A}$ of degree n and $\mathfrak{A} = \mathfrak{B}_1 \times \mathfrak{A}_2$ where $\mathfrak{A}_2$ is simple of degree n_2 over its center $\mathfrak{F}$, $n = 2n_2$. If $\mu > 1$ we have $m \geqq 6$ and define $G_4G_k = H_{k-3}$ for $k = 5, \cdots, m$ and see that $G_2H_j - H_jG_2 = G_3H_j - H_jG_3 = 0$ for $j = 2, \cdots, m-3$. Also $H_j^2 = -\beta_4\beta_{j+3}I$, $H_iH_j = -H_jH_i$ for $i \neq j$ and $i, j = 2, \cdots, m-3$. But then the H_i are in $\mathfrak{A}_2$ and we have the same situation as above with m replaced by $m - 3$, n by n_2, $\mathfrak{A}$ by $\mathfrak{A}_2$. It follows that 2 divides n_2, 2^2 divides n. After μ such steps we obtain the result desired. Moreover if $n = m$ we have shown that 2^μ divides $n = 3\mu$, $3\mu + 1$, or $3\mu + 2$. However if $\mu > 3$ we have $2^\mu > 3\mu + 2$, if $\mu = 3$ then $2^\mu = 8$ does not divide 9, 10, 11, and if $\mu = 2$ then $2^2 = 4$ only divides $3\mu + 2 = 8$. If $\mu = 1$ then $2^\mu = 2$ only divides $3\mu + 1 = 4$ and if $\mu = 0$ then $n = 1, 2$.

where D_1 is the non-singular r-rowed diagonal matrix with diagonal elements $\alpha_1 = 1, \alpha_2, \cdots, \alpha_r$, D_2 is the r-rowed diagonal matrix with $\alpha_{r+1}, \cdots, \alpha_{2r}$ as diagonal elements, D is the $(n - 2r)$-rowed non-singular diagonal matrix with diagonal elements $\alpha_{2r+1}, \cdots, \alpha_n$. We now prove the

LEMMA 5. *The equation* (1) *is possible only if there exist linear forms* $u_{2r+1}, \cdots, u_n$ *in* $y_1, \cdots, y_m$ *such that*

$$g(y) = \alpha_{2r+1}^{-1}(\alpha_{2r+1}u_{2r+1}^2 + \cdots + \alpha_n u_n^2). \tag{76}$$

For proof we observe first that $g(y)[A + A'] = G_y(A + A')G_y'$. Write

$$G_y = \begin{pmatrix} G & H \\ K & L \end{pmatrix}, \tag{77}$$

where G is a $2r$-rowed square matrix and the number of rows and columns in H, K, L are thereby determined. Then

$$A + A' = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}, \qquad C = \begin{pmatrix} 0 & I_r \\ I_r & 0 \end{pmatrix}. \tag{78}$$

But then C is non-singular and

$$\begin{pmatrix} g(y)C & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} G & H \\ K & L \end{pmatrix}\begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} G' & K' \\ H' & L' \end{pmatrix} = \begin{pmatrix} GCG' & GCK' \\ KCG' & KCK' \end{pmatrix}. \tag{79}$$

Hence[19] $g(y)C = GCG'$, G is non-singular, $KCG' = 0$, $K = 0$. Also

$$G_y A(G_y)' = G_y\begin{pmatrix} A_1 & 0 \\ 0 & D \end{pmatrix}\begin{pmatrix} G' & 0 \\ H' & L' \end{pmatrix} = \begin{pmatrix} G & H \\ 0 & L \end{pmatrix}\begin{pmatrix} A_1G' & 0 \\ DH' & DL' \end{pmatrix}. \tag{80}$$

But the diagonal elements of $g(y)A - G_yA(G_y)'$ are all zero and thus the diagonal elements of $g(y)D - LDL'$ are zero. Write $L = (u_{ij})$ for $i, j = 1, \cdots, q$, where $q = n - 2r$ and the u_{ij} are linear forms in $y_1, \cdots, y_m$. Compute the element in the first row and column of this matrix and obtain

$$\alpha_{2r+1}g(y) - (\alpha_{2r+1}u_{11}^2 + \cdots + \alpha_n u_{1q}^2) = 0.$$

This completes our proof.

A quadratic form $g(y) = yBy'$ is a diagonal form $\beta_1 y_1^2 + \cdots + \beta_m y_m^2$ if and only if $B + B' = 0$. Then every quadratic form equivalent to $g(y)$ is also diagonal. Now if $u_i = \sum_{j=1}^{m} \lambda_{ij}y_j$ with the λ_{ij} in $\mathfrak{F}$ we have $u_i^2 = \lambda_{i1}^2 y_1^2 + \cdots + \lambda_{im}^2 y_m^2$. By (76) we have

---

[19] Observe that then we are led to the study of the same type of equation as in the non-singular case. We show below that $g(y)$ is diagonal and later that there is then an associated field $\mathfrak{K}$ which is purely inseparable of degree 2^t and exponent two over $\mathfrak{F}$. But then the set of all $2r$-rowed square matrices contains a subfield generated by the matrices G_i equivalent to $\mathfrak{K}$ and thus 2^t divides $2r$.

LEMMA 6. *The equation* (1) *is possible for* $n > 2r$ *only if* $g(y)$ *is diagonal.*

It follows immediately that if $n > 2r$ for a quadratic form $f(x)$ with $r > 0$ it cannot permit composition.

8. Diagonal quadratic forms

The theory of (1) for $f(x)$ diagonal and $\mathfrak{F}$ of characteristic two is very simple and the results obtainable with little difficulty. We write $f(x)$ as in (11) with the α_i all not zero. Define $\mathfrak{K}(f)$ to be the purely inseparable field $\mathfrak{F}(u_1, \cdots, u_n)$, $u_i^2 = \alpha_i$, so that $\mathfrak{K}(f)$ has degree 2^t and exponent one or two over $\mathfrak{F}$, $n \leqq 2^t$. Define

$$u(x) = x_1u_1 + \cdots + x_nu_n, \tag{81}$$

and let $\mathfrak{L}(f)$ be the set of all values $u(\xi)$ for $\xi = (\xi_1, \cdots, \xi_n)$ and the ξ_i in $\mathfrak{F}$. Clearly $\mathfrak{L}(f)$ is a linear subspace of order $\nu \leqq n$, $\nu \leqq 2^t$, of $\mathfrak{K}(f)$. Moreover $f(x) = [u(x)]^2$.

By permuting the indeterminates $x_1, \cdots, x_n$ if necessary we may assume that $u_1, \cdots, u_\nu$ are linearly independent in $\mathfrak{F}$ and that $u_k = \sum_{j=1}^{\nu} \alpha_{kj}u_j$ for $k = \nu + 1, \cdots, n$. Then the linear transformation

$$x_j = w_j - \sum_{k=\nu+1}^{n} \alpha_{kj}w_k, \qquad x_k = w_k \qquad (j = 1, \cdots, \nu;\ k = \nu + 1, \cdots, n) \tag{82}$$

is non-singular and carries $f(x)$ into $\alpha_1w_1^2 + \cdots + \alpha_\nu w_\nu^2 + (\sum \alpha_{kj}^2\alpha_j + \alpha_k)w_k^2$. But $u_k^2 = \alpha_k = \sum \alpha_{kj}^2\alpha_j$. Hence (82) carries $f(x)$ into

$$\alpha_1w_1^2 + \cdots + \alpha_\nu w_\nu^2. \tag{83}$$

We have thus proved

LEMMA 7. *Let* $f(x)$ *be a diagonal quadratic form and* $\mathfrak{L}(f)$ *have order* ν *over* $\mathfrak{F}$. *Then* $f(x)$ *is equivalent to a form* $\alpha_1x_1^2 + \cdots + \alpha_\nu x_\nu^2$.

Note that ν is an invariant of $f(x)$ and that $f(x)$ is not equivalent to a diagonal form in fewer than ν variables. We now prove

LEMMA 8. *Let* $f(x) = \alpha_1x_1^2 + \cdots + \alpha_nx_n^2$, $h(x) = \gamma_1x_1^2 + \cdots + \gamma_mx_m^2$ *such that* n *and* m *are the respective orders of* $\mathfrak{L}(f)$, $\mathfrak{L}(h)$. *Then there exist linear forms* $w_1, \cdots, w_n$ *in* $x_1, \cdots, x_m$ *such that* $h(x) = f(w)$ *if and only if* $m \leqq n$, $\mathfrak{L}(h)$ *is contained in* $\mathfrak{L}(f)$. *Also then* $f(x)$ *and* $h(x)$ *are equivalent if and only if* $\mathfrak{L}(h) = \mathfrak{L}(f)$.

For we may construct a field $\mathfrak{F}(u_1, \cdots, u_n, v_1, \cdots, v_m) = \mathfrak{W}$ such that $u_i^2 = \alpha_i$, $v_j^2 = \gamma_j$. Then $\mathfrak{W}$ contains both $\mathfrak{L}(f)$ and $\mathfrak{L}(h)$, $f(x) = [u(x)]^2$, $h(x) = [v(x)]^2$, where $u(x)$ is defined by (81) and $v(x) = v_1x_1 + \cdots + v_mx_m$. Also $h(x) = f(w)$ implies that $v(x) = u(w)$, every quantity of $\mathfrak{L}(h)$ is in $\mathfrak{L}(f)$, $\mathfrak{L}(f)$ contains $\mathfrak{L}(h)$. Conversely if $\mathfrak{L}(f)$ contains $\mathfrak{L}(h)$ we may take $v_1, \cdots, v_m$ as part of a basis of $\mathfrak{L}(f)$ and then replace $f(x)$ by an equivalent form with $\alpha_i = \gamma_i$ for $i = 1, \cdots, m$. Then $h(x) = f(w)$ where $w_i = x_i$ for $i = 1, \cdots, m$ and $w_k = 0$ for $k = m + 1, \cdots, n$. The last part of our lemma is a trivial consequence of what we have already proved.

The result above implies that we may take $n - 2r$ to be the order of the matrix of $\alpha_{2r+1}x_{2r+1}^2 + \cdots + \alpha_n x_n^2$ in Lemma 5 and obtain $m \leqq n - 2r$ in that lemma.

We now prove

LEMMA 9. *Let $f(x)$ be a diagonal form $x_1^2 + \alpha_2 x_2^2 + \cdots + \alpha_n x_n^2$ where n is the order of $\mathfrak{L}(f)$. Then $f(x)g(y) = f(z)$ as in (1), (2), if and only if $g(y)$ is equivalent to a diagonal quadratic form $\beta_1 y_1^2 + \cdots + \beta_m y_m^2$ such that $\mathfrak{L}(f)$ contains $\mathfrak{K}(g)$. Moreover n is divisible by the order 2^t of $\mathfrak{K}(g)$ over $\mathfrak{F}$.*

The result above states that $\mathfrak{L}(f)$ is a linear space of order ν over $\mathfrak{K}(g)$, $n = 2^t\nu$,

$$\mathfrak{L}(f) = \mathfrak{K}(g) + \mathfrak{K}(g)d_2 + \cdots + \mathfrak{K}_g d_\nu . \tag{84}$$

Then we may take a basis of $\mathfrak{K}(g)$ of order $\sigma = 2^t$ over $\mathfrak{F}$ to be a set of quantities $v_1, \cdots, v_\sigma$ such that $v_i^2 = \beta_i$ in $\mathfrak{F}$ and also take $d_j^2 = \gamma_j$ for $j = 2, \cdots, \nu$. Then we may take $g(y) = \beta_1 y_1^2 + \cdots + \beta_m y_m^2$ and $f(x) = \sum_{i=1}^{\sigma} \beta_i x_i^2 + \sum_{i=1}^{\sigma} \beta_i\gamma_2 x_{\sigma+i}^2 + \cdots + \sum_{i=1}^{\sigma} \beta_i\gamma_\nu x_{(\nu-1)\sigma+i}^2$.

To prove this result we note that $f(e) = 1$ and thus $g(y) = f(w)$ for $w = (w_1, \cdots, w_n)$ and the w_i linear forms in $y_1, \cdots, y_m$. Thus $g(y)$ is a diagonal form. Then $g(y)$ is a diagonal form such that $\mathfrak{L}(g)$ is contained in $\mathfrak{L}(f)$. But then $g(y) = [v(y)]^2$ for $v(y) = v_1y_1 + \cdots + v_my_m$ and $v_i^2 = \beta_i$ where the v_i are a basis of $\mathfrak{L}(g)$. If a and b are in $\mathfrak{L}(g)$ the quantity a is in $\mathfrak{L}(f)$. However $[u(x)]^2[v(y)]^2 = [u(z)]^2$, $u(x)v(y) = u(z)$ and ab is in $\mathfrak{L}(f)$. It follows that the product of any two quantities of $\mathfrak{L}(g)$ is in $\mathfrak{L}(f)$, and $\mathfrak{L}(f)$ contains the ring of all polynomials in the quantities of $\mathfrak{L}(g)$. This is the field $\mathfrak{K}(g)$. Conversely if $\mathfrak{L}(f)$ contains $\mathfrak{K}(g)$ we know that $\mathfrak{L}(f) = [\mathfrak{K}(g)]\mathfrak{B}$ where $\mathfrak{B} = (1, d_2, \cdots, d_\nu)$ over $\mathfrak{F}$ and $1, d_2, \cdots, d_\nu$ are a basis of $\mathfrak{L}(f)$ over $\mathfrak{K}(g)$. Then $\mathfrak{L}(g)\mathfrak{K}(g)$ is in $\mathfrak{K}(g)$, $\mathfrak{L}(g)\mathfrak{L}(f)$ is contained in $\mathfrak{L}(f)$. However the equation $g(x)f(y) = f(z)$ is precisely equivalent to the statement that $\mathfrak{L}(g)\mathfrak{L}(f)$ is contained in $\mathfrak{L}(f)$.

If $g(y)$ is equivalent to $f(x)$ we may require that $m = n$ is the order of $\mathfrak{L}(f)$. Then $\alpha_1 f(y) = f(w)$ and $f(y) = \alpha_1^{-1} f(w)$. Lemma 8 then states that $f(x)$ is equivalent to $\alpha_1^{-1} f(x)$. It follows that we may take $\alpha_1 = 1$. Then we may apply Lemma 9 to obtain the result that $\mathfrak{L}(f)$ contains $\mathfrak{K}(f)$. Then $\mathfrak{L}(f) = \mathfrak{K}(f)$. Thus *a diagonal quadratic form over a field of characteristic two permits composition if and only if it is the norm form of a purely inseparable field of degree 2^t and exponent two over $\mathfrak{F}$.* We have thus completed the proof of our theorem.

THE UNIVERSITY OF CHICAGO

ANNALS OF MATHEMATICS
Vol. 43, No. 1, January, 1942

ON INTEGRAL GEOMETRY IN KLEIN SPACES

BY SHIING-SHEN CHERN

(Received September 3, 1940)

The classical results in integral geometry found by Crofton, Poincaré, Cartan, and recently developed by Blaschke[1] and his school, are mostly restricted to Euclidean spaces. It is the object of this paper to give the fundamental concepts of integral geometry in a general space of Klein, by which we mean a number space of n dimensions with a transitive r-parameter group G_r of transformations. The discussion is mainly based on Cartan's theory of Lie's groups.[2]

The paper is divided into three sections. In §1 we give a brief summary of Cartan's theory of Lie's groups with some results concerning the measures of geometrical elements. In §2 we define the incidence of the geometrical elements of different fields, which plays a fundamental rôle in all subsequent discussions. As applications of the general notions we give in the last section two formulas which are respectively generalizations of the well-known formulas of Crofton and Cauchy.

1. Some Fundamental Notions in Klein Spaces

Let G_r be an abstract r-parameter group of Lie[3] with the parameters $a^1, \cdots, a^r$, so that G_r denotes an r-dimensional space and $a^1, \cdots, a^r$ are the coordinates in the space. If S_a denotes a point in G_r with the coordinates $a^1, \cdots, a^r$, there are defined in G_r two simply-transitive groups of transformations

$$S_a \to S_c S_a\,, \qquad S_a \to S_a S_c\,, \tag{1}$$

called respectively the first and the second groups of parameters. There exists one, and only one, set of r linearly independent Pfaffian forms

$$\omega^1(a, da), \cdots, \omega^r(a, da)$$

invariant under the first group of parameters. They are determined up to a linear transformation with constant coefficients and have the property that any

[1] Cf. W. Blaschke, *Vorlesungen über Integralgeometrie*, Bd. I Leipzig 1936; Bd. II Leipzig 1937. These books will be cited respectively as I. G. I and I. G. II.

[2] Cf. E. Cartan, *La théorie des groupes finis et continus et la géométrie différentielle traitée par la méthode du repère mobile*, Paris 1937. This book will be cited as Cartan, Théorie des groupes.

[3] For the definition of an abstract group of Lie we may follow that given by Cartan in his book "*La théorie des groupes finis et continus et l'analysis situs*," Mémorial des Sciences Mathématiques, Fasc. XLII, Paris 1930.

exterior differential form[4] of degree p invariant under the first group of parameters is of the form

$$\sum_{i_1,\cdots,i_p} A_{i_1\cdots i_p}[\omega^{i_1}\cdots\omega^{i_p}], \tag{2}$$

where $A_{i_1\cdots i_p}$ are constants. According to Cartan, we call $\omega^1, \cdots, \omega^r$ the relative components of G_r. They satisfy the equations of structure of Maurer-Cartan

$$(\omega^i)' = \sum_{j,k=1}^{r} c^i_{jk}[\omega^j\omega^k], \qquad c^i_{jk} + c^i_{kj} = 0, \qquad i = 1, \cdots, r, \tag{3}$$

where c^i_{jk} are the constants of structure of G_r.

Now let g be a subgroup of G_r. This subgroup g and its left-hand cosets Sg fill up the whole space G_r and have the property that no two of them can coincide without being identical. Thus the varieties Sg are the integral varieties of a completely integrable Pfaffian system. Owing to the invariance of the totality of the cosets Sg with respect to the first group of parameters, the left-hand members of the Pfaffian system are linear combinations of $\omega^1, \cdots, \omega^r$ with constant coefficients. We may suppose the system to be

$$\omega^1 = 0, \cdots, \omega^n = 0. \tag{4}$$

The theorem of Frobenius then gives

$$c^i_{jk} = 0, \qquad i = 1, \cdots, n; \qquad j, k = n+1, \cdots, r, \tag{5}$$

which are the necessary and sufficient conditions for the system (4) to be completely integrable. We say that the subgroup g defines a field of geometrical elements such that each element of the field is represented by a left-hand coset of g or by an integral variety of (4).

We may interpret the geometrical elements defined by (4) as the points of a space E. The first group of parameters becomes then a group of transformations in E, so that, with some necessary assumptions on the analyticity of the equations of the transformations, E is a space in the sense of Klein. It is clear that any space of Klein can be obtained this way. In fact, we need only take G_r to be the group of transformations in the space and g the subgroup of G_r leaving invariant a fixed point.

Let $x^1(a^1, \cdots, a^r), \cdots, x^n(a^1, \cdots, a^r)$ be n independent first integrals of (4), so that $dx^1, \cdots, dx^n$ are n linearly independent combinations of $\omega^1, \cdots, \omega^n$. Then we have

$$[dx^1 \cdots dx^n] = \Delta(a^1, \cdots, a^r)[\omega^1 \cdots \omega^n], \qquad \Delta \neq 0. \tag{6}$$

By the measure of a domain D of elements in E we shall mean an n-tuple integral of the form

$$J = \int_D f(x^1, \cdots, x^n)[dx^1 \cdots dx^n] \tag{7}$$

---

[4] For the notions of exterior differential forms and exterior derivation cf. E. Cartan, *Leçons sur les invariants intégraux*, Chap. VI, VII, Paris 1922.

extended over D such that its value is invariant under the group of transformations in E, i.e., under the first group of parameters. Since this property holds for all domains D, we see from (6) that it is expressed by

$$f\Delta = \text{constant.}$$

Hence *a necessary and sufficient condition for the elements of E to possess a measure is that the function $\Delta(a)$ in* (6) *be a function of $x^1, \cdots, x^n$ only.* The measure is then defined up to a constant factor and is given by

$$J = \int_D [\omega^1 \cdots \omega^n]. \tag{8}$$

The condition for the existence of a measure can also be expressed in terms of the constants of structure. Let d and δ be two operations such that d denotes a displacement in the set of elements in E and δ a displacement on an integral variety of (4). The condition for the existence of measure can be written as

$$\delta[\omega^1 \cdots \omega^n] = 0.$$

But we have

$$\omega^1(\delta) = \cdots = \omega^n(\delta) = 0.$$

The equations of structure (3) then give

$$\delta\omega^i(d) = 2\sum_{k=1}^{n}\sum_{j=n+1}^{r} c^i_{jk}\omega^j(\delta)\omega^k(d), \qquad i = 1, \cdots, n.$$

From the last equations we get

$$\delta[\omega^1 \cdots \omega^n] = 2\sum_{j=n+1}^{r}\sum_{i=1}^{n} c^i_{ji}\omega^j(\delta)[\omega^1 \cdots \omega^n].$$

Therefore *a necessary and sufficient condition for the elements defined by* (4) *to possess a measure is*

$$\sum_{i=1}^{n} c^i_{ji} = 0, \qquad j = n+1, \cdots, r. \tag{9}$$

2. Definition of the Incidence of Elements of Different Fields

Consider two fields of geometrical elements, defined respectively by (4) and

$$w^1 = 0, \cdots, w^m = 0, \tag{10}$$

where w^i $(i = 1, \cdots, m)$ are linear combinations of the relative components with constant coefficients. Each of the systems (4) and (10) being completely integrable, the same must be true of the combined system

$$\omega^1 = 0, \cdots, \omega^n = 0, \qquad w^1 = 0, \cdots, w^m = 0. \tag{11}$$

Of the last system suppose s of them ($s > m$, $s > n$) be linearly independent, so that there exist $m + n - s$ relations of the form

$$\sum_{i=1}^{n} b_{\lambda i}\omega^i + \sum_{j=1}^{m} c_{\lambda j}w^j = 0, \qquad \lambda = 1, \cdots, m + n - s, \tag{12}$$

where $b_{\lambda i}$, $c_{\lambda j}$ are constants. The integral varieties of (11) are of dimension $r - s$. Denoting the integral varieties of (4), (10), (11) respectively by V_{r-n}, V_{r-m}, V_{r-s}, we see that *a V_{r-s} lies completely on a V_{r-n} (or V_{r-m}) if and only if it has a point in common with V_{r-n} (or V_{r-m}). It follows that a V_{r-n} and a V_{r-m} have a V_{r-s} in common if and only if they have a point in common.*

By virtue of the above properties we define *an element N of* (4) *and an element M of* (10) *to be incident when their corresponding integral varieties V_{r-n}, V_{r-m} have a V_{r-s} in common.*

Consider the elements M incident with a given element N. In the group space G_r each of the integral varieties V_{r-m} corresponding to M cuts V_{r-n} in a V_{r-s}. The totality of V_{r-s} on a fixed V_{r-n} depends on $s - n$ parameters. Therefore *the elements of the field* (10) *incident with a given element N of* (4) *depend on $s - n$ parameters.* The same property remains naturally true when the two fields are interchanged.

It is important to notice that this definition of incidence includes the ordinary notions of incidence in Euclidean geometry, affine geometry, projective geometry, etc. as particular cases. Take, for instance, the group of motions in the Euclidean plane

$$\begin{cases} x^* = x\cos c - y\sin c + a, \\ y^* = x\sin c + y\cos c + b, \end{cases} \tag{13}$$

where a, b, c are parameters. The relative components are

$$\begin{cases} \omega^1 = \cos c\,da + \sin c\,db, \\ \omega^2 = -\sin c\,da + \cos c\,db, \\ \omega^3 = dc. \end{cases} \tag{14}$$

As the equations of points and straight lines we may take respectively

$$\omega^1 = 0, \qquad \omega^2 = 0 \tag{15}$$

and

$$\omega^1 = 0, \qquad \omega^3 = 0. \tag{16}$$

In the three-dimensional group space (a, b, c) the elements of the field (15) are represented by lines parallel to the c-axis and those of (16), by the lines

$$\cos c\cdot a + \sin c\cdot b = \text{const.}, \qquad c = \text{const.}$$

By interpreting a, b as the Cartesian coordinates in the plane and the equation

$$\cos c\cdot a + \sin c\cdot b = \text{const.}$$

as the equation of line in Hesse's normal form, we see immediately that our definition of incidence coincides with the notion of incidence in the ordinary sense.

Before concluding this section, we want to remark that, instead of using the integral varieties in the group space to characterize the elements of a field, we may employ the geometrically more intuitive idea of families of frames (in French "repère").[5] But it is sufficient for our purpose to restrict ourselves to one point of view.

3. The Generalized Crofton's Formula and Cauchy's Formula

Let us first consider the geometry in the Euclidean plane. It is well known that the measures for points and lines exist and that they are given respectively by[6]

$$\int [\omega^1 \omega^2] \qquad \text{and} \qquad \int [\omega^1 \omega^3].$$

Crofton's formula in its simplest form asserts that the measure of the lines incident with the points of a curve is equal to a constant multiple of the length of the curve, each line being counted as many times as the number of its points of intersection with the curve.

To generalize this formula, we must first have the generalized notion of the length of a curve for a p-dimensional variety of points under an arbitrary group of transformations of Lie. For this purpose, take a p-dimensional variety V_p formed by the elements of the field (4). Let $u^1, \cdots, u^p$ be the parameters on V_p. Then $\omega^1, \cdots, \omega^n$ are linear combinations of $du^1, \cdots, du^p$ on V_p. On eliminating $du^1, \cdots, du^p$, we may express $\omega^{p+1}, \cdots, \omega^n$ as linear combinations of $\omega^1, \cdots, \omega^p$ in the form

$$\omega^\alpha = \sum_{k=1}^{p} \xi_k^\alpha \omega^k, \qquad \alpha = p+1, \cdots, n,$$

where ξ_k^α are functions of the parameters $a^1, \cdots, a^r$ of the group and of $u^1, \cdots, u^p$. By the method of moving frames of Cartan, it is in general possible (by supposing the variety V_p to be sufficiently general) to determine some or all of the ξ_k^α to be constants or functions of the parameters $u^1, \cdots, u^p$ such that the determination is invariant under transformations of the group G_r. If, up to a certain step in the determination of ξ_k^α, the exterior differential forms of degree p

$$[\omega^{i_1} \cdots \omega^{i_p}],$$

where $i_1, \cdots, i_p$ run over all combinations of $1, \cdots, n$, depend on $u^1, \cdots, u^p$ only and differ from each other only by constant factors, we say that the variety V_p possesses a *p-dimensional area*, which is equal to the integral of any one of the

[5] Cf. Cartan, *Théorie des groupes*, Chap. V, or H. Weyl, *The Classical Groups*, Princeton 1939, pp. 16–17.

[6] Blaschke, I. G. I, pp. 5–7.

differential forms and is thus determined up to a constant factor. It is of course possible that the area of V_p does not exist, as in the case that V_p is a quadric in projective space. But in the cases which usually occur to us, the p-dimensional area of V_p under G_r exists.

In the case $p = 1$ the p-dimensional area so defined leads to the Pick's invariant of a curve,[7] which includes the affine arc, projective arc, etc. as particular cases. Pick has proved that *if a group of transformations of order r in the xy-plane transforms transitively the elements of contact of order $r - 2$*

$$x, y, y' = \frac{dy}{dx}, \cdots, y^{(r-2)} = \frac{d^{r-2}y}{dx^{r-2}},$$

then a plane curve possesses an intrinsic parameter invariant under the group considered. This intrinsic parameter is called the invariant of Pick. Its existence is an easy consequence of our preceding discussions. In fact, it is only necessary to take $x, y, y', \cdots, y^{(r-2)}$ to be parameters of the group and

$$\omega^1 = A^1(x, y, \cdots, y^{(r-2)})\,dx + B^1(x, y, \cdots, y^{(r-2)})\,dy = 0,$$
$$\omega^2 = A^2(x, y, \cdots, y^{(r-2)})\,dx + B^2(x, y, \cdots, y^{(r-2)})\,dy = 0$$

to be the equations of points. Along a curve we have $dy = y'dx$, so that ω^2 is a constant multiple of ω^1, and the integral

$$\int \omega^1$$

gives the intrinsic parameter of the curve.

Another important particular case of the notion of p-dimensional area is the case that $p = s - m$ and that V_p consists of all the elements of the field (4) incident with a fixed element of the field (10). In this case, the relations between $\omega^1, \cdots, \omega^n$ are obtained from (12) by setting $w^1 = \cdots = w^m = 0$ and are therefore

$$\sum_{i=1}^{n} b_{\lambda i}\omega^i = 0, \qquad \lambda = 1, \cdots, m + n - s.$$

Since $b_{\lambda i}$ are constants, we may take as the p-dimensional area of this V_p the integral

$$\int [\omega^1 \cdots \omega^p].$$

If the value of this integral over V_p is finite, we call it *the measure of the elements of* (4) *about a fixed element of* (10).

With these preparations we can give the generalized Crofton's formula in the following theorem:

---

[7] Cf., for example, G. Kowalewski, *Allgemeine Natürliche Geometrie und Liesche Transformationsgruppen*, Berlin 1931, pp. 106–110.

Let two fields of elements M, N be defined respectively by the equations (10), (4). *Let $p = m + n - s$ and let V_p be a variety of p dimensions formed by the elements of M. If the measure of the elements of N incident with the elements of V_p and the p-dimensional area F of V_p both exist and are finite, then*

$$\int [\omega^1 \cdots \omega^n] = cF, \qquad c = \text{constant}, \tag{17}$$

where the integral is extended over the elements of N incident with the elements of V_p, each element being counted as many times as the number of elements of V_p with which it is incident.

To prove this theorem, consider the $p = m + n - s$ relations (12). Since the relations are independent with respect to $\omega^1, \cdots, \omega^n$, we may solve them in terms of $\omega^1, \cdots, \omega^p$, obtaining

$$\omega^\lambda = \sum_{j=p+1}^{n} e_j^\lambda \omega^j + \sum_{k=1}^{m} f_k^\lambda w^k, \qquad \lambda = 1, \cdots, p,$$

where e_j^λ, f_k^λ are constants. Setting

$$\bar{\omega}^\lambda = \omega^\lambda - \sum_{j=p+1}^{n} e_j^\lambda \omega^j, \qquad \lambda = 1, \cdots, m + n - s,$$

and denoting $\bar{\omega}^\lambda$ again by ω^λ, we have

$$\omega^\lambda = \sum_{k=1}^{m} f_k^\lambda w^k, \qquad \lambda = 1, \cdots, m + n - s. \tag{18}$$

The proof of our theorem then depends on a new form of the expression $[\omega^1 \cdots \omega^n]$. We apply first the method of moving frames of Cartan to V_p. Since by hypothesis the p-dimensional area of V_p exists we can attach to each element of V_p a frame such that within the p-parameter family of frames so attached the exterior differential forms

$$[w^{i_1} \cdots w^{i_p}]$$

where $i_1, \cdots, i_p$ run over all combinations of $1, \cdots, m$, differ from each other only by constant factors. In the group space $G_r V_p$ is represented by a p-parameter family of left-handed cosets with respect to a subgroup h. To the frames attached to the elements of V_p there corresponds in G_r a p-dimensional variety W_p such that on every coset corresponding to an element of V_p there lies one and only one point of W_p. Now every integral variety Sh of (10) is cut by the integral varieties of (4) incident to it in the $(r - s)$-dimensional integral varieties V_{r-s} of (11). The totality of V_{r-s} on Sh depends on $s - m$ parameters. To define a system of coordinates for the V_{r-s} on Sh we may first set up a system of coordinates $\lambda^1, \cdots, \lambda^{s-m}$ for the V_{r-s} on h. By the transformation $S_a \to SS_a$ (S being fixed) the variety h is carried to Sh and we define the coordinates of a V_{r-s} on h to be the coordinates of its image on Sh.

Let $u^1, \cdots, u^p$ be the parameters on V_p. As the coordinates of an element

of the field N incident with an element of V_p we may take $u^1, \cdots, u^p, \lambda^1, \cdots, \lambda^{s-m}$. Now the expression $[\omega^1 \cdots \omega^n]$ depends only on the elements of the field N under consideration. This property, called by Blaschke the property of invariance of choice ("Wahlinvarianz"), may be interpreted as follows: Choose on each integral variety of (4) a point and compute the relative components $\omega^1, \cdots, \omega^n$ on the n-dimensional variety so obtained. The resulting expression is independent of the choice of these points.

This being clear, we may chose the points on the integral varieties of (4) to be on the V_{r-s} of the integral varieties corresponding to the elements of V_p. Its totality forms a variety of $p + s - m = n$ dimensions, which we call V_n. To find the relative components on V_n notice that the relative components $\omega^i(a, da)$ $(i = 1, \cdots, r)$ are the parameters of the infinitesimal transformation $S_a^{-1}S_{a+da}$ and that by choosing the infinitesimal transformations X_if $(i = 1, \cdots, r)$ such that $X_{m+1}f, \cdots, X_rf$ generate the subgroup h leaving invariant a fixed element 0 of the field M, the infinitesimal transformation $S_a^{-1}S_{a+da}$ carrying 0 to a neighboring element P is of the form

$$w^1X_1f + \cdots + w^mX_mf + \omega^{m+1}X_{m+1}f + \cdots + \omega^rX_rf,$$

where $w^1, \cdots, w^m$ are exactly the left-hand members of the system (10).

Let $\theta^1, \cdots, \theta^m$ denote the relative components $w^1, \cdots, w^m$ on W_p, $\bar{\omega}^{p+1}, \cdots, \bar{\omega}^n$ the relative components $\omega^{p+1}, \cdots, \omega^n$ on Sh, and let $w^1, \cdots, w^m, \omega^{p+1}, \cdots, \omega^n$ denote these components on V_n. As the independent components on V_n we may then take p independent forms among the $\theta^1, \cdots, \theta^m$, and $\bar{\omega}^{p+1}, \cdots, \bar{\omega}^n$, the latter being Pfaffian forms in $\lambda^1, \cdots, \lambda^{s-m}$ only. Now every point of V_n is of the form

$$S_a = S_bR,$$

where S_b and R belong respectively to W_p and h. By writing

$$S_{a+da} = S_{b+db}R^1, \qquad R^1 \in h,$$

we get

$$S_a^{-1}S_{a+da} = R^{-1}S_b^{-1}S_{b+db}R \cdot R^{-1}R^1.$$

Since $R^{-1}R^1$ belongs to h, it produces no effect on 0 The relative components $w^1, \cdots, w^m$ are thus transformed according to the transformations of the adjoint group which correspond to transformations of h. As R transforms between themselves the infinitesimal transformations of h, we have

$$(19) \qquad \begin{cases} w^1 = a_1^1\theta^1 + \cdots + a_m^1\theta^m, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ w^m = a_1^m\theta^1 + \cdots + a_m^m\theta^m, \end{cases}$$

where $a_k^i (i, k = 1, \cdots, m)$ are functions of $\lambda^1, \cdots, \lambda^{s-m}$.

On the other hand, for the components $\omega^{p+1}, \cdots, \omega^n$ on V_n we must have

$$(20)\qquad \begin{cases} \omega^{m+n-s+1} = \bar{\omega}^{m+n-s+1} + b_1^{m+n-s+1}\theta^1 + \cdots + b_m^{m+n-s+1}\theta^m, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \omega^n = \bar{\omega}^n + b_1^n\theta^1 + \cdots + b_m^n\theta^m, \end{cases}$$

since $\omega^i (i = p+1, \cdots, n)$ reduce to $\bar{\omega}^i$ when $u^1, \cdots, u^p$ are constants. Thus we get

$$[\omega^{p+1} \cdots \omega^n] \equiv [\bar{\omega}^{p+1} \cdots \bar{\omega}^n], \qquad (\text{mod } \theta^1, \cdots, \theta^m).$$

Finally, from (18) and (19), we find

$$(21)\qquad [\omega^1 \cdots \omega^n] = f(\lambda^1, \cdots, \lambda^{s-m})[\bar{\omega}^{p+1} \cdots \bar{\omega}^n d\Omega],$$

where $d\Omega$ denotes the element of the p-dimensional area of V_p.

It is easy to finish the proof of our theorem from the last formula. In fact, evaluating the integral

$$\int f(\lambda^1, \cdots, \lambda^{s-m})[\bar{\omega}^{p+1} \cdots \bar{\omega}^n\, d\Omega]$$

by first holding an element of V_p fixed, we get a constant which is the same for all elements of V_p. Integration over V_p then gives the p-dimensional area of V_p.

From this proof we observe that even when the measure of the elements of N incident with the elements of V_p becomes infinite, there still exists a generalized formula of Crofton. In fact, we may make the following restriction: On each variety of $s - m$ dimensions V_{s-m} consisting of the elements of (4) incident with a fixed element of (10) take a fixed domain D such that the integral

$$\int f(\lambda^1, \cdots, \lambda^{s-m})[\bar{\omega}^{p+1} \cdots \bar{\omega}^n]$$

is finite and constant for all elements of (10). Then the measure of the elements of (4) incident with the elements of V_p and belonging to the domain D at each element of incidence is finite and will be given by Crofton's formula (17).

We now come to generalize the formula of Cauchy. Consider a surface S in Euclidean space and the planes E intersecting the surface. If $\dot{E}$ denotes the density of the planes and U_E the perimeter of the curve of intersection of the surface by the plane, then the formula of Cauchy is[8]

$$(22)\qquad \int U_E \dot{E} = \frac{\pi^2}{2} F,$$

where F denotes the area of S and the integral in the left-hand side is extended over all planes intersecting S.

To give a generalization of this formula take a variety V_p of p dimensions formed by the elements of the field M. As in the proof of Crofton's formula (17)

[8] Blaschke, I. G. II, p. 73.

let $u^1, \cdots, u^p$ be the parameters on V_p and let us attach to each element of V_p a frame. Almost all the arguments used above are valid, including the formulas (19), (20). The only difference is that the variety, formally denoted by V_n, is now of $p + s - m$ dimensions. This variety V will be cut by an integral variety V_{r-n} of (4) in a variety $V_{p+s-m-n}$ of $p + s - m - n$ dimensions (supposing $p > m + n - s$). Through each point of $V_{p+s-m-n}$ there passes an integral variety of (10). Its totality consists of all the elements of V_p incident with V_{r-n}. To find the independent components on $V_{p+s-m-n}$ it is only necessary to set $\omega^\lambda = 0$ in (18). The independent Pfaffian forms in the equations so obtained

$$\sum_{k=1}^{m} f_k^\lambda w^k = 0, \qquad \lambda = 1, \cdots, m + n - s$$

then give the independent components on $V_{p+s-m-n}$. The generalized Cauchy's formula states that *the integral over all elements of* (4) *incident with* V_p *of a* $(p + s - m - n)$*-dimensional integral invariant of* $V_{p+s-m-n}$ *of the form*

$$\int \sum C_{i_1 \cdots i_{p+s-m-n}} [w^{i_1} \cdots w^{i_{p+s-m-n}}], \tag{23}$$

where $C_{i_1 \cdots i_{p+s-m-n}}$ *are functions of* $\lambda^1, \cdots, \lambda^{s-m}$ *only, is equal to a constant multiple of the p-dimensional area of* V_p, *provided that both quantities are finite.*

It is only necessary to prove the formula for the case when the sum (23) contains one term. Suppose it be

$$\int C[w^1 \cdots w^{p+s-m-n}].$$

By making use of (18), (20), we get

$$C[w^1 \cdots w^{p+s-m-n} \omega^1 \cdots \omega^n] = F(\lambda^1, \cdots, \lambda^{s-m})[\theta^1 \cdots \theta^p \bar{\omega}^{m+n-s+1} \cdots \bar{\omega}^n],$$

where F is a function of $\lambda^1, \cdots, \lambda^{s-m}$. The generalized Cauchy's formula as stated above is then an easy consequence of the relation obtained.

It may be helpful to give an example of the above general discussions. Consider the three-dimensional Euclidean space with its group of motions. Instead of introducing the group space we may take a fixed right-hand rectangular trihedral T_0 in space and take all such trihedrals T as the elements of the space. Then the motions in space and the trihedrals T are in one-to-one correspondence such that to T corresponds the motion carrying T_0 to T. As the parameters of the group we may take the parameters of T. Let P be the origin of T and $\vec{e}_1, \vec{e}_2, \vec{e}_3$ the three unit vectors along the axes of T. Then the relative components of the group are defined by the equations

$$\begin{cases} dP = \omega^1 \vec{e}_1 + \omega^2 \vec{e}_2 + \omega^3 \vec{e}_3, \\ d\vec{e}_i = \omega_i^1 \vec{e}_1 + \omega_i^2 \vec{e}_2 + \omega_i^3 \vec{e}_3, \end{cases} \qquad \omega_j^i + \omega = 0, \qquad (i, j = 1, 2, 3). \tag{24}$$

The equations of structure of the group are

$$(25)\qquad \begin{cases} (\omega^i)' = \sum_{k=1}^{3} [\omega^k \omega_k^i], \\ (\omega_k^i)' = \sum_{j=1}^{3} [\omega_k^j \omega_j^i], \end{cases} \qquad i, k = 1, 2, 3.$$

As the coset in the group space corresponding to a point we take the family of trihedrals having the point as origin. As the coset corresponding to a straight line we take the trihedrals whose origin is on the line and whose vector $\vec{e}_3$ is along the line. Finally, the coset corresponding to a plane is formed by the trihedrals with origin on the plane and with the third unit vector $\vec{e}_3$ perpendicular to the plane. By these definitions, the equations of the points, lines, and planes are respectively

$$(26)\qquad \begin{cases} \omega^1 = 0, & \omega^2 = 0, & \omega^3 = 0; \\ \omega^1 = 0, & \omega^2 = 0, & \omega_3^1 = 0, & \omega_3^2 = 0; \\ \omega^3 = 0, & \omega_3^1 = 0, & \omega_3^2 = 0. \end{cases}$$

For all these three fields of elements the measures exist, as may be verified by using the criterion (9). They are then given by the integrals

$$(27)\qquad \int [\omega^1 \omega^2 \omega^3], \qquad \int [\omega^1 \omega^2 \omega_3^1 \omega_3^2], \qquad \int [\omega^3 \omega_3^1 \omega_3^2].$$

Consider now the fields of points and lines and denote them by M, N respectively. Then

$$m = 3, \qquad n = 4, \qquad s = 5.$$

Put

$$w^1 = \omega^1, \qquad w^2 = \omega^2, \qquad w^3 = \omega^3.$$

The relations (18) become

$$\omega^1 = w^1, \qquad \omega^2 = w^2.$$

To obtain the formula of Crofton in this case take a surface Σ and attach to each point of the surface a trihedral $P\vec{v}_1\vec{v}_2\vec{v}_3$ with origin at this point and with $\vec{v}_3$ normal to the surface. Choose the trihedral $P\vec{e}_1\vec{e}_2\vec{e}_3$ attached to a line through P such that $\vec{e}_1$ is on the plane $P\vec{v}_1\vec{v}_2$. Then we have

$$(28)\qquad \begin{cases} \vec{v}_1 = -\sin\psi\,\vec{e}_1 - \cos\psi\cos\varphi\,\vec{e}_2 + \cos\psi\sin\varphi\,\vec{e}_3, \\ \vec{v}_2 = \cos\psi\,\vec{e}_1 - \sin\psi\cos\varphi\,\vec{e}_2 + \sin\psi\sin\varphi\,\vec{e}_3, \\ \vec{v}_3 = \sin\varphi\,\vec{e}_2 + \cos\varphi\,\vec{e}_3 \end{cases}$$

where φ, ψ are the parameters of the lines through P, φ being the angle between $\vec{e}_3$ and $\vec{v}_3$. From

$$dP = \theta^1\vec{v}_1 + \theta^2\vec{v}_2 = w^1\vec{e}_1 + w^2\vec{e}_2 + w^3\vec{e}_3,$$

we get

$$\begin{aligned} \omega^1 &= -\theta^1 \sin\psi + \theta^2 \cos\psi, \\ \omega^2 &= -\theta^1 \cos\psi \cos\varphi - \theta^2 \sin\psi \cos\varphi. \end{aligned} \tag{29}$$

It follows that

$$[\omega^1\omega^2\omega_3^1\omega_3^2] = \cos\varphi[\bar{\omega}_3^1\bar{\omega}_3^2\theta^1\theta^2]. \tag{30}$$

This is a particular case of the formula (21). The classical proof of Crofton's formula is based on this relation.

Next, take the fields of points and planes and denote them by M, N respectively. Then

$$m = 3, \qquad n = 3, \qquad s = 5.$$

Put

$$w^1 = \omega^1, \qquad w^2 = \omega^2, \qquad w^3 = \omega^3.$$

The set of relations (18) consists only of one equation

$$\omega^3 = w^3.$$

As by (29), we get

$$w^3 = \theta^1 \cos\psi \sin\varphi + \theta^2 \sin\psi \sin\varphi. \tag{31}$$

Along the curve of intersection of the plane $P\vec{e}_1\vec{e}_2$ with Σ, we have

$$w^3 = 0$$

and

$$dP = w^1\vec{e}_1$$

so that $\vec{e}_1$ is the tangent and w^1 the element of arc. Then we get

$$[w^1\omega^3\omega_3^1\omega_3^2] = [w^1w^3\bar{\omega}_3^1\bar{\omega}_3^2] = -\sin\varphi[\theta^1\theta^2\bar{\omega}_3^1\bar{\omega}_3^2]. \tag{32}$$

From this relation we derive easily Cauchy's formula.

TSING HUA UNIVERSITY
KUNMING, CHINA